

2009

## Predictive modeling techniques with application to the Cerulean Warbler (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region

Matthew Buhrl Shumar  
*West Virginia University*

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Shumar, Matthew Buhrl, "Predictive modeling techniques with application to the Cerulean Warbler (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region" (2009). *Graduate Theses, Dissertations, and Problem Reports*. 2785.  
<https://researchrepository.wvu.edu/etd/2785>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

**Predictive modeling techniques with application to the Cerulean Warbler  
(*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region**

**Matthew Buhrl Shumar**

**Thesis submitted to the Davis College of Agriculture, Forestry, and Consumer  
Sciences at West Virginia University in partial fulfillment of the requirements  
for the degree of**

**Master of Science  
in  
Wildlife and Fisheries Resources**

**Petra Bohall Wood, Ph.D., Chair  
Michael P. Strager, Ph.D.  
Wayne E. Thogmartin, Ph.D.**

**Division of Forestry and Natural Resources**

**Morgantown, West Virginia  
2009**

**Keywords: Bayesian, cerulean warbler, classification and regression trees,  
count data, *Dendroica cerulea*, Markov chain Monte Carlo, North American  
Breeding Bird Survey, resolution, scale, spatial modeling**

## ABSTRACT

### **Predictive modeling techniques with application to the Cerulean Warbler (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region**

**Matthew Buhrl Shumar**

Many statistical approaches have been used for developing predictive models for wildlife presence/absence and abundance, each with varying levels of accuracy and complexity. As concerns for declining species intensify and anthropogenic impacts on habitats increase, the ability to quickly quantify and map species distributions and abundances over large regions will become increasingly important. To date, there is no set of best practices for modeling specific wildlife groups. My primary objectives with this thesis were to 1) compare model techniques for ease of use and accuracy, and 2) compare resolution of species occurrence data and its effect on model accuracy.

For the first objective, I compared two modeling techniques that range from moderately quick and simplistic (decision trees) to conceptually and computationally complex (hierarchical spatial models). I used North American Breeding Bird Survey counts with a suite of explanatory variables to predict presence and abundance of cerulean warblers (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region. Of the decision tree methods, cerulean warbler occurrence was most accurately described by presence/absence models. Regression tree abundance models under-predicted counts and had low accuracy. Hierarchical spatial models predicted abundance of cerulean warblers similar to actual counts, and with better overall accuracy than regression trees. All techniques produced models using similar variables; interior forest and percent forest were most important for identifying areas with cerulean warblers.

For the second objective, I compared two model types, differing in the resolution of the species distribution data. I used North American Breeding Bird Survey (NABBS) counts with a suite of explanatory variables to predict presence and abundance of cerulean warblers (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region (BCR28). Decision trees were created for route-level and stop-level analyses of presence and abundance. Additionally, output maps have typically been resolved to the resolution of the environmental spatial datasets with little attention given to the scale at which the predictions represent. Using the modeling results, predictive distribution maps were created for cerulean warblers with appropriate resolutions for each model group. Route-level decision trees performed better than stop-level models for predicting both presence and abundance of cerulean warblers. Similar to raw NABBS distribution data, cerulean warblers were predicted to occur in highest concentrations in the central portions of the BCR. Poor performance of stop-level models may result from a mismatch of resolution of environmental data to species survey data, or lack of important environmental covariates at the stop-level scale. The results of this study highlight the importance of correctly matching the resolution of the species distribution data to the resolution of environmental covariates and the extent of analysis.

The results and relationships highlighted in this thesis may serve to direct management and monitoring for the cerulean warbler, and other migratory passerines.

## ACKNOWLEDGMENTS

Funding for this project was granted by the USFWS Science Support Partnership. Logistic and administrative support was provided by the USGS West Virginia Cooperative Fish and Wildlife Research Unit and the Division of Forestry and Natural Resources at West Virginia University.

Many people helped with this project, for which I am incredibly grateful. First, I would like to thank my advisor, Dr. Petra Wood who gave me the opportunity to work on a graduate assistantship at WVU. I could not have asked for a better advisor. Her patience and guidance throughout this entire project have been invaluable. My graduate committee members Dr. Michael Strager and Dr. Wayne Thogmartin provided continual support, guidance and suggestions about my thesis. Thank you, Wayne for your patient assistance in guiding me into the world of Bayesian statistics.

Special thanks to Patrick McElhone, Alison Mynsberge, and Jackie Strager who provided an immense amount of assistance and support throughout this project. I thank other fellow graduate students who provided advice, encouragement, good birding company, and a few laughs along the way. Additionally, I would like to thank Becky Nestor for coordinating everything under the sun, Brian Smith for his help throughout this project, and all the North American Breeding Bird Survey volunteers who took the time to collect spatial coordinates used in this project.

I am grateful to my family for all their love, support and encouragement. Finally, I would like to thank my wife, Molly McDermott, who in so many ways made the completion of this project possible. Thank you.

## TABLE OF CONTENTS

Abstract .....	ii
Acknowledgements .....	iii
List of Tables .....	vi
List of Figures .....	vii
List of Appendices .....	x

### **Chapter 1: Literature review of predictive modeling on the Cerulean Warbler**

Introduction .....	2
Landscape change in the Appalachian Mountains .....	5
Cerulean Warblers in the Appalachian .....	7
Predictive modeling and songbirds .....	9
Literature Cited .....	13

### **Chapter 2: Comparison of classification and regression trees and Bayesian modeling approaches for predicting the occurrence of cerulean warblers**

Abstract .....	24
Introduction .....	25
Study Area .....	27
Methods .....	28
Data sets .....	28
Decision tree models .....	29
Hierarchical spatial models .....	30
Model evaluation .....	31
Mapping .....	32
Results .....	32
Decision trees .....	32
Hierarchical spatial models .....	33
Discussion .....	33
Management Implications .....	36
Literature Cited .....	36

### **Chapter 3: Comparison of Breeding Bird Survey route- and stop-level spatial models with application to cerulean warblers**

Abstract .....	57
Introduction .....	58
Methods .....	60
Study area .....	60
Data sets .....	60
Decision tree models .....	62
Model evaluation .....	63
Mapping .....	63
Results .....	63
Route-level models .....	63
Stop-level models .....	64

Discussion .....	65
Literature Cited .....	67

#### **Chapter 4: Technological limitations of regional species modeling**

Introduction .....	87
Software Limitations .....	87
Derivation of landscape metrics .....	88
Data formatting for modeling .....	89
The modeling process .....	89
Hardware Limitations .....	90
Modeling and Mapping Limitations .....	91
Technology and Research Needs .....	92
Literature Cited .....	93

## LIST OF TABLES

### Chapter 1:

Table 1. Key variables from previous Cerulean Warbler modeling efforts.....	21
---	----

### Chapter 2:

Table 1. Variables used to predict presence/absence and abundance of cerulean warblers in the Appalachian Bird Conservation Region (BCR28). Variables were only used at scales that would be biologically relevant to the focal species.....	42
--	----

Table 2. Confusion matrix from the classification tree for cerulean warbler presence/absence. 104 (randomly chosen) NABBS routes were withheld from model creation for evaluation.....	42
--	----

Table 3. Model averaged posterior distributions of standardized model parameters with 95% credibility intervals. All estimates were included in the creation of a final distribution map for cerulean warbler abundance.....	43
--	----

Table 4. DIC values and model weights for final models at each scale for hierarchical spatial analyses.....	43
---	----

Table 5. Standard deviation associated with random effects. Larger standard deviation values represent larger variation explained by the random effect.....	44
---	----

### Chapter 3

Table 1. Variables used to predict presence/absence and abundance of cerulean warblers in the Appalachian Bird Conservation Region (BCR28). Variables were only used at scales that would be biologically relevant to the focal species. For models created at each scale, B = both resolution groups and S = stop-level models only.....	70
---	----

Table 2. Confusion matrix from the route-level classification tree for cerulean warbler presence/absence. Randomly chosen NABBS routes (n=35) were withheld from model creation for evaluation.....	71
---	----

Table 3. Confusion matrix from the stop-level classification tree for cerulean warbler presence/absence. Randomly chosen NABBS stops (n=1607) were withheld from model creation for evaluation.....	71
---	----

## LIST OF FIGURES

### Chapter 1:

Figure 1. Breeding range of Cerulean Warbler as derived from an interpolation of Breeding Bird Survey counts (Sauer et al. 2007).....22

### Chapter 2:

Figure 1. Breeding range of the cerulean warbler as derived from an interpolation of Breeding Bird Survey (NABBS) counts (Sauer et al. 2008). Black triangles represent centroids for NABBS routes where at least one Cerulean Warbler was detected between 1998 and 2004. Grey dots represent routes where no Cerulean Warblers were detected during the time period.....45

Figure 2. Partners in Flight physiographic regions and study area boundary.....46

Figure 3. Neighborhood structure for NABBS routes within the Appalachian Mountains Bird Conservation Region (BCR28). Only first order neighbors were used in the adjacency matrix to model the spatial structure of counts. Neighborhood structure was created by tessellating North American Breeding Bird survey routes (NABBS; n=348) within BCR28.....47

Figure 4. Classification tree model for cerulean warbler presence (PRES)/absence (ABS). Values below each node represent the number of routes assigned to each associated category...48

Figure 5. Predicted presence/absence of cerulean warblers within the Appalachian Bird Conservation region. Results are based on classification tree analysis and represent sum of cerulean warblers detected on a NABBS survey route (n=348 routes).....49

Figure 6. Regression tree model for cerulean warbler abundance. Each node ends with the estimated abundance of cerulean warblers and the number of routes classified by the node.....50

Figure 7. Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on regression tree analysis and represent sum of cerulean warblers detected on a NABBS survey route (n=348 routes).....51

Figure 8. Observed maximum North American Breeding Bird Survey (NABBS) counts (a: n = 104 validation routes) and observed yearly NABBS counts (b: n=594) during 1998 – 2004 for cerulean warblers in the Appalachian Bird Conservation Region (BCR28) compared to expected (predicted) counts from the best regression tree and hierarchical spatial model of cerulean warbler abundance. Observed counts were withheld from model construction. Dashed lines represent a one-to-one correspondence between observed and predicted counts. We fitted observed versus expected counts with simple linear regression (solid line). Each point may represent more than one route or stop.....52



Figure 9. Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on a hierarchical spatial analysis and represent sum of cerulean warblers detected on a NABBS survey route (n=348 routes).....53

Figure 10. Variation between predicted abundance maps created by decision trees and hierarchical spatial models. Green cells indicate areas where cerulean warblers are predicted in higher abundances by hierarchical spatial models. Red cells indicate areas where cerulean warblers are predicted in higher abundances by decision tree models. Maps were created individually using results from statistical models and compared by assessing the mathematical difference between raster cells in ArcGIS 9.2.....54

Figure 11. Mapped route effects (n=348 NABBS routes) for the 300 m hierarchical spatial model. Positive numbers indicate an increasing effect of route on abundance, while negative numbers indicate a decreasing effect of route on abundance.....55

### Chapter 3:

Figure 1. Breeding range of the cerulean warbler as derived from an interpolation of Breeding Bird Survey (NABBS) counts (Sauer et al. 2008). Black triangles represent centroids for NABBS routes where at least one Cerulean Warbler was detected between 1998 and 2004. Grey dots represent routes where no Cerulean Warblers were detected during the time period.....72

Figure 2. Partners in Flight physiographic regions and study area boundary.....73

Figure 3. Route-level classification tree model for cerulean warbler presence (PRES)/absence (ABS). Values below each node represent the number of routes assigned to each associated category.....74

Figure 4. Predicted presence/absence of cerulean warblers within the Appalachian Bird Conservation region. Results are based on route-level classification tree analysis (n=115 NABBS routes).....75

Figure 5. Route-level regression tree model for cerulean warbler abundance. Each node ends with the estimated abundance of cerulean warblers and the number of routes classified by the node.....76

Figure 6. Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on route-level regression tree analysis. Abundance is defined by the sum of counts on a NABBS survey route (n=115 routes).....77

Figure 7. Observed mean North American Breeding Bird Survey (NABBS) counts (a: n = 35 validation routes and b: n=1607 validation stops) during 1998 – 2004 for cerulean warblers in the Appalachian Bird Conservation Region (BCR28) compared to expected (predicted) counts from the best regression tree. Observed counts were withheld from model construction. Dashed lines represent a one-to-one correspondence between observed and predicted counts. We fitted

observed versus expected counts with simple linear regression (solid line). Each point may represent more than one route or stop.....78

Figure 8. Stop-level classification tree model for cerulean warbler presence (PRES)/absence (ABS). Values below each node represent the number of routes assigned to each associated category.....79

Figure 9. Predicted presence/absence of cerulean warblers within the Appalachian Bird Conservation region. Results are based on stop-level classification tree analysis (n=5,355 NABBS stops).....80

Figure 10. Stop-level regression tree model for cerulean warbler abundance. Each node ends with the estimated abundance of cerulean warblers and the number of routes classified by the node.....81

Figure 11. Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on stop-level regression tree analysis (n=5,355 NABBS stops).....82

Figure 12. Variation between route- and stop-level maps for cerulean warbler presence/absence. Green cells indicate areas predicted to contain cerulean warblers by route- but not stop-level models. Red cells indicate areas predicted to contain cerulean warblers by stop- but not route-level models. Maps were created individually using classification trees and compared in ArcGIS 9.2.....83

Figure 13. Variation between route- and stop-level maps for cerulean warbler relative abundance. Green cells indicate areas where cerulean warblers are predicted in higher abundances by route-level models. Red cells indicate areas where cerulean warblers are predicted in higher abundances by stop-level models. Maps were created individually using regression trees and compared by assessing the mathematical difference between raster cells in ArcGIS 9.2. Route-level maps were standardized by dividing abundances by 50 (number of NABBS stops on a route) for direct comparison to stop-level maps.....84

Figure 14. Map resolution comparison between (a) route-level regression tree results and (b) stop-level regression tree results. Area represents the boundaries of Pennsylvania, West Virginia, and Ohio. Cell size for the route-level map was resampled to a cell size of 1566.5 m x 1566.5 m to represent the listening radius for cerulean warblers over 50 NABBS stops. Cell size for the stop-level map was 221.5 m x 221.5 m to represent the listening radius for cerulean warblers over a single NABBS stop.....85

## Chapter 4:

Figure 1. Examples of ‘trap’ and ‘black-box’ errors produced by WinBUGS during the model iteration.....95

Figure 2. Breeding Bird Survey stop locations within the Ohio Hills Partners in Flight physiographic region (PIF22).....96

## LIST OF APPENDICES

### Chapter 4:

Appendix 1. Complete process from data collection to model results for a classification and regression tree (CART) analysis.....	97
Appendix 2. Complete process from data collection to model results for a hierarchical spatial analysis using WinBUGS.....	98

# **Chapter 1**

## **Literature review of predictive modeling on the Cerulean Warbler**

## INTRODUCTION

Landscapes throughout the Appalachians have been in constant change because of anthropogenic disturbances over the last several hundred years. Changes such as forest fragmentation and loss of habitat at both local and regional scales have been implicated as important factors contributing to decreases in bird populations (O'Conner et al. 1996, Sauer et al. 2008). This is likely to increase the probability of local extinctions of some species and will lead to a reduction in forest species richness (Boulinier et al. 1998).

Urban sprawl coupled with an ever-expanding human population may be one of the biggest threats to wildlife. With mid-range estimates of 8-10 billion people on earth by the end of the 21<sup>st</sup> century, loss of habitat and direct conflict with humanity will be a problem that wildlife biologists and land managers will increasingly have to face (Maurer 1996, Lutz and Qiang 2002). For example, Stein et al. (2005) identified the watersheds projected to have the largest increase in housing density. Among those was the Little Kanawha watershed in West Virginia, which lies in the midst of the core range of the Cerulean Warbler (*Dendroica cerulea*), a songbird species of concern that has faced steep declines in recent decades. Additionally, as human populations increase, demand for fuels and resources will also rise. Throughout the Appalachian Mountains, coal, natural gas, and timber products are all important industries. Extraction of these resources creates temporal changes in vegetation that may displace or eliminate avian communities (Maurer et al. 1981, King and DeGraaf 2000, Bosworth 2003).

Most management and land use practices have come under scrutiny with declines in avian populations. Although direct loss of habitat is one of the most obvious detrimental effects to wildlife, other related factors can complicate management issues. Large-scale activities such as mountaintop removal mining-valley fill (MTRVF) can alter the spatial configuration of forested habitats, creating edge and area effects, thereby negatively affecting forest-dwelling songbirds (Wood et al. 2006a). Mature-forest dependent songbird species such as the Cerulean Warbler negatively respond to large-scale land practices (Bosworth 2003, Weakland and Wood 2005, Wood et al. 2006a). Other taxa such as reptiles, amphibians, insects, and fungi may show similar or more severe negative responses (Martel and Mauffette 1997, Johnson et al. 2003).

Also, the Brown-headed Cowbird (*Molothrus ater*), a brood parasite that negatively affects reproductive success of forest songbirds, has benefited from these landscape changes and expanded its range throughout portions of the eastern United States (Robinson et al. 1995). This

may be due to conversion of forested land to agriculture and increases in forest fragmentation (Brittingham and Temple 1983). Deforestation in the late nineteenth century aided this conversion (Bonan 1999, Ramankutty and Foley 1999).

Forest interior bird species, formerly isolated from cowbirds because of a lack of grassland habitat and corridors, have become exposed to cowbird parasitism as forest fragmentation has increased edge and provided access for cowbirds (Brittingham and Temple 1983). Though the actual percentage of nests parasitized is low for some forest species, increases in parasitism have driven already rare species such as the Kirtland's Warbler (*Dendroica kirtlandii*) to critically low population levels (Lowther 1993). However, cowbird populations have been declining in the east (Sauer et al. 2008), possibly because of the return of the land to a forested state and loss of farmland. LaDeau et al. (2007) suggested that West Nile Virus may also be implicated in songbird declines. Compounding factors make it difficult to isolate specific causes of most avian population declines.

Although regional-scale removal of trees as a result of timber harvesting and mining may negatively affect forest interior species, these activities benefit early-successional and grassland species by providing habitat. Contour mines provide temporary breeding habitat for Golden-winged Warblers (*Vermivora chrysoptera*; Bulluck and Buehler 2006), while large-scale mining has provided areas suitable for Henslow's Sparrows (*Ammodramus henslowii*; Bajema and Lima 2001) and Grasshopper Sparrows (*A. savannarum*; Ammer 2003). Early-successional species have recently received much attention because of drastic population declines (Litvaitis 1993, Hunter et al. 2001, Sauer et al. 2008). This attention has been used as justification for timber and mining companies whose industries defend their actions as creating early-successional habitat through the process of resource extraction. While this is true, area sensitive mature forest songbirds may be harmed as a result.

Biologists and land managers must prioritize which specific lands and habitat types need to be restored, protected, and/or managed to most effectively achieve conservation objectives for the species of interest (Will et al. 2005, Fitzgerald et al. 2008, Thogmartin et al. 2008). Each species is uniquely impacted by various land use practices, and management plans must be designed accordingly. However, research is often disconnected from land management. Collaborations must be formed so that scientists understand the needs of managers, and the best information is applied to management. Large-scale resource planning for species of

conservation concern like the Cerulean Warbler requires an interdisciplinary approach, specialized knowledge, open communication, and integrated thinking (Zabel et al. 2002).

Predictive modeling, combined with Geographic Information Systems (GIS), is becoming a common tool for use in wildlife management. Many modeling efforts are aimed at managing for multiple species using insight gained from a model of single-species habitat relations. While multi-species management is a desirable goal, models must be tailored to fit the requirements of individual species to be most effective. To date many avian models have had relatively poor predictability rates because of problems with differences in species' habitat occurrence, issues of scale, sampling biases, lack of knowledge concerning species' life histories, and many other factors (Beard et al. 1999, Gutzwiller and Barrow 2001, Wood et al. 2006b). With advances in GIS, remote sensing, and additional knowledge about species' life histories, models can be developed that will more accurately predict species occurrences and relative abundances. Additionally, models with high predictive success at one geographic location may not be successful or appropriate at other locations or scales (Heglund 2002). For example, models may be created for a population that may be at or near carrying capacity and then erroneously applied to areas that are under-populated relative to carrying capacity, or vice versa (Boone and Krohn 2002).

The Cerulean Warbler was chosen as the focal species for this modeling work because it is a species of high conservation concern. Its core range lies within the Appalachian Mountains Bird Conservation Region (BCR28; Figure 1). Using spatial datasets for Cerulean Warbler locations, environmental characteristics, and layers representing threats to the species, predictive models were created to aid in the prioritization of landscapes for conservation. Robbins et al. (1992) identified six chief constraints to Cerulean Warblers on the breeding ground: (1) loss of mature deciduous forest, especially along stream valleys; (2) fragmentation and increasing isolation of remaining mature deciduous forest; (3) change to shorter rotation periods and even-aged management so that less deciduous forest habitat reaches maturity; (4) environmental degradation from acid rain and stream pollution; (5) loss of key tree species, especially oaks (*Quercus* spp.) from oak wilt (*Ceratocystis fagacearum*) and gypsy moths (*Lymantria dispar*), sycamores (*Platanus occidentalis*) from a fungus (*Gnomonia leptostyla*), elms (*Ulmus* spp.) from Dutch elm disease (*Ophiostoma ulmi*), and American chestnuts (*Castanea dentata*) from chestnut blight (*Cryphonectria parasitica*); and (6) brood parasitism by the Brown-headed Cowbird.

Forest and other landcover variables representing these threats were included in modeling. Brown-headed Cowbirds are excluded from the model, because they do not pose a serious problem in the Appalachian portion of the Cerulean Warbler's breeding range (Robinson et al. 1995, Duguay et al. 2001, W. Thogmartin unpublished data).

This thesis is organized into four chapters. Chapter 1 includes an introduction to the modeling study and the relevant literature. Chapters 2 and 3 will present and discuss two case studies with application to the Cerulean Warbler. Chapter 2 compares classification and regression tree (CART) and Bayesian modeling approaches. Chapter 3 is a comparison of Breeding Bird Survey route-level and stop-level spatial models. The final chapter discusses complications with the modeling processes and suggests directions for future research.

## **LANDSCAPE CHANGE IN THE APPALACHIAN MOUNTAINS**

Forests throughout the eastern United States are in a state of constant change because of urbanization and continuing resource extraction. Some changes are temporary, while others such as sprawl result in permanent loss of habitat. Changes in forest patch size, type, and seral stage can all impact whether a species occupies a particular area (Maurer et al. 1981, King and DeGraaf 2000, Bosworth 2003). Additionally, major changes in local population sizes of individuals can have drastic effects on the overall species' population. Individuals may not settle into an area that is lacking conspecifics, which may ultimately keep local populations low or cause local extinctions (Muller et al. 1997). Cerulean Warblers may exhibit a degree of coloniality during the breeding season as a result of conspecific attraction (COSEWIC 2003).

Large-scale activities such as surface mining alter the spatial configuration of forested habitats, creating edge and area effects that negatively affect forest songbirds (Wood et al. 2006a). Mining techniques such as mountaintop removal/valley fill can impact areas on the order of 2,000+ ha in size, converting a landscape that was predominately forested to one that is principally early-successional (Wood et al. 2001). These large-scale mining techniques often change the structure of the landscape in the Appalachians, resulting in nutrient pollution to aquatic systems, a decrease in atmospheric moisture, and decreased forest habitat (Wickham et al. 2006). Additionally, because of the fragmentation and creation of edge, loss of interior forest is as much as five times higher than the direct forest loss attributable to mountaintop mining (Wickham et al. 2006). The United State Environmental Protection Agency estimated that the



4,856,247 ha Southern Appalachian region was 92% forest, and that mountaintop mining will remove 6.8% of the forest between 1992 and 2012 (US EPA 2005).

Several forest-dependent bird species of high conservation priority, such as the Cerulean Warbler, Louisiana Waterthrush (*Seiurus motacilla*), Worm-eating Warbler (*Helmitheros vermivorus*), Wood Thrush (*Hylocichla mustelina*), Acadian Flycatcher (*Empidonax virescens*), and Scarlet Tanager (*Piranga olivacea*), have the core of their breeding ranges in forested uplands and headwater streams (Wood et al 2006b). These areas often overlap with industrial land uses like surface mining. Several studies (Bosworth 2003, Weakland and Wood 2005, Wood et al. 2006a) have found negative effects on Cerulean Warbler abundance and territory density as a result of these landscape-level impacts.

Timber harvesting occurs throughout the Appalachian Mountains and is a major industry in some regions. Different silvicultural prescriptions result in mixed responses from different taxa and responses from the public. While clearcutting is viewed negatively by the general public, the resulting early-successional habitat is critical for species such as Golden-winged Warblers (Hunter et al. 2001). Other types of harvesting are being examined for their effects on wildlife. Two-aged harvesting, which leaves residual trees for a deferred removal, is currently being used in the central Appalachians as an alternative to clearcutting (Smith et al. 1989). Two-age cuts may function similarly to clearcuts, while being more visually acceptable to the general public (Miller et al. 2006). Baker and Lacki (1997) and McDermott and Wood (2009) demonstrated that the level of harvest (clearcut, and two-age treatments with varying amount of residuals) had no discernible effect on the bird communities (i.e., birds responded similarly to clearcuts and two-aged harvests). In contrast, Boardman and Yahner (1999) and Duguay et al. (2001) found two-age cuts to be more beneficial. Even-aged stands that retain some overstory trees (a feature of two-age cuts) have a greater diversity of breeding birds compared to similarly aged stands without residual trees. Differences in results among the studies may be explained by differences in geographic location. Therefore, modeling results may be difficult to extrapolate because of different responses in other geographic regions.

Landscape distribution and intensity of timber harvesting are also important when modeling and managing for a particular species, although large-scale spatial datasets are difficult to obtain. Fearer (2007) used Forest Inventory Analysis (FIA) datasets with North American Breeding Bird Survey (NABBS) data to explain variations in occupancy and abundance for

Cerulean Warblers and other forest bird species across spatial and temporal scales. However, one of the major limitations with FIA data is that precise coordinates are not available to the general public. Lack of precise spatial coordinates forces an interpolation of the data and may limit analyses to a courser scale.

## **CERULEAN WARBLERS IN THE APPALACHIANS**

The Appalachian Bird Conservation Region (BCR28) provides habitat for 234 species of breeding, migrant, and wintering birds (Atlantic Coast Joint Venture 2005). This diverse group of avian species is a consequence of the longitudinal range, diverse habitats, and elevations from 0-1,981 m. The Appalachian Mountain region also contains many areas that are undergoing rapid human development. Eastern West Virginia, Virginia, North Carolina, and other areas are undergoing drastic land changes to accommodate human population growth and lifestyle changes (Loveland and Acevedo 2006).

The Cerulean Warbler is a small (11.5 cm, 8-10 g) Neotropical migrant that winters in the Andes Mountains (Hamel 2000). During the breeding season, the Cerulean Warbler can be found throughout portions of the eastern United States and parts of the Midwest (Figure 1) in mature deciduous forest. The central Appalachians represent the core of the Cerulean Warbler's breeding range (Figure 1) and are believed to be a stronghold for breeding populations (Rosenberg et al. 2000). The core of its range overlaps with areas altered by the coal mining technique of mountaintop removal/valley fill and other types of surface mines. Much of the remaining forested area in the MTRVF region is mixed-mesophytic hardwood forest (approximately 58%), which is a forest type that is unique to the Central Appalachians. In addition to land use practices for resource extraction, parts of the Cerulean Warbler's breeding range overlap with many expanding urban centers. As cities expand and freeway systems allow urban residents quick access to ever-shrinking natural areas, wildlife populations will be impacted negatively. Corridor H, a current freeway project that will connect the Carolinas with the District of Columbia, will bisect the Monongahela National Forest in West Virginia (West Virginia Division of Highways 2003), eliminating habitat in areas with known breeding populations of Cerulean Warblers.

The Cerulean Warbler has experienced a rangewide decline averaging approximately 4.1% per year, which amounts to a population reduction of 70% since 1966 (Sauer et al. 2008).

Thogmartin (2006) projected an 83% chance that the population will experience a 90% decrease in 100 years. Consequently, the Cerulean Warbler is considered a species of special concern over its breeding range and was recently considered for listing as “threatened” under the Endangered Species Act. Although it was not listed (Federal Register 2006), the Cerulean Warbler is still considered a focal species of research and management priority by the U.S. Fish and Wildlife Service and other agencies. In response to growing concern over the steep declines, the Cerulean Warbler Technical Group was formed to help identify research needs and priority management actions (Hamel et al. 2004). Along with the creation of the Cerulean Warbler Technical Group, a large study has been initiated by several universities to obtain information on the effects of silvicultural practices on breeding populations.

Several variables have been identified from prior research as important to Cerulean Warbler habitat use and reproductive success. Canopy cover, canopy gaps (or amount of canopy closure), distance from large-scale edge, slope position, and aspect have been suggested as the best predictors of Cerulean Warbler density (Jones et al. 2001, Bosworth 2003, Weakland and Wood 2005, Buehler et al. 2006, Perkins 2006, Wood et al. 2006a, Wood 2006b). Different aspects contain diverse microhabitat features, and there is likely some micro-component that this species uses. Northeastern and eastern slopes tend to be more biologically productive due to ideal solar exposure and moisture levels (Beers et al. 1966) and may be of particular importance to Cerulean Warblers (Hartman 2006, Wood et al. 2006a).

Because the Cerulean Warbler is a forest interior species, patch size of core-forest and amount of forest in the landscape are other key habitat requirements. Though this species tends to avoid large-scale edge, often created by industrial land use practices and urban development (Wood et al. 2006a), small breaks in the forest canopy seem to be used more than or in proportion to their availability (Weakland and Wood 2005, Perkins 2006). Currently, there is no moderate-resolution spatial product available to represent canopy heterogeneity for the entire Appalachian Mountains BCR. However, canopy closure and structure may be correlated with other datasets that are more readily available. Thogmartin (personal communication) found canopy closure to be highly correlated with proportion of forest for the Central Hardwoods BCR.

The Cerulean Warbler has shown different patterns in habitat use across its breeding range, using riparian bottomlands in the southern extent of the range and ridge-top areas in the core breeding areas (Hamel 2000, Rosenberg et al. 2000, Weakland and Wood 2005, Wood et al.

2006b). In the western extent of the range, Thogmartin et al. (2004) indicated that Cerulean Warblers were most abundant in drier upland areas embedded within a larger matrix of forested wetlands. Other species have also shown a distinct dichotomy for habitat preference across their breeding range. Black-and-White Warblers have been associated with mature forested areas; however in the Allegheny Mountains of West Virginia, Black-and-White Warblers also use young harvested forest extensively (McDermott and Wood 2009). This illustrates the value of identifying local variables important to habitat use for each species of interest. Accordingly, variables that are preferred at different scales across different portions of birds' breeding ranges need to be accounted for in predictive models.

## **PREDICTIVE MODELING AND SONGBIRDS**

Predictive modeling for birds, particularly migratory passerines, is a complicated task because these species require separate breeding, migratory stopover, and wintering habitats. Requirements among species are often quite variable, which makes multi-species management approaches difficult. Integrating present sources of information into models is further complicated by the nature of the data, as most large-scale surveys collect indices of population size (e.g., relative abundance) rather than unbiased estimates of population size (Thogmartin et al. 2006).

Many modeling methods poorly accommodate both observer and spatial effects because modeling these spatially autocorrelated counts is not practical using standard statistical approaches (Thogmartin et al. 2004). Variability in the skills of NABBS observers can be considerable, and there is evidence for a temporal component to this variability (Link and Sauer 2002). The temporal variability is partly explained by the year in which a particular observer begins to survey routes, with improvement and consistency coming with experience. Failure to accommodate for observer effects introduces substantial bias in trend estimation (Sauer et al. 1994).

In addition to observer biases, position and arrangement of count locations can also bias results, often presenting overly optimistic or pessimistic estimates, depending on region and species of interest (Bibby et al. 1992, Keller and Scallan 1999, Peterjohn 2001). Furthermore, there is an underlying assumption that observations equate to patterns (Li and Wu 2004), yet in some cases, this may be an artifact of the scale of analysis (McPherson et al. 2006). Changes in

scale may result in a change in observed patterns (Levin 1992, Riitters et al. 1997, McPherson et al. 2006). Patterns observed at a particular scale may be the result of processes operating at another scale or the combination of a suite of scales (McPherson et al. 2006).

Scale, both temporal and spatial, is an important component in ecological research, and must be taken into account when possible. There are three major components that characterize wildlife habitat relationships with concern to spatial scale: 1) The resolution of the species' distribution data; 2) the resolution of the habitat variables; and 3) the extent of the study area (Tobalske 2002). The latter two are the major components of a proper definition of scale. The terms scale and landscape have been used in different ways by a variety of professional fields (e.g., ecology, remote sensing) with conflicting meanings. It is important to clarify the term, particularly since the line between ecology, modeling, and remote sensing is often blurred. A typical definition of scale is loosely centered around the concept of size. Turner et al. (1989) defined scale as the total size of an area being considered, independent of the resolution (or time period) that is measured or represented. Landscape ecologists have expanded this definition to include two independent and unrelated elements: "grain," defined as resolution, and "extent," defined as total size (Hutson 2002). In addition to identifying appropriate scales for analysis, models must also be able to minimize Type I (errors of commission) and Type II (errors of omission) errors.

Besides quantifying error, it is important to interpret error results correctly. This is particularly true with errors of commission. Error reported may be *actual* error (the species is not present on the site) or *apparent* error (the species is present on the site but has not been recorded as a result of incomplete field inventories; Schaefer and Krohn 2002).

There are many statistical approaches for model creation, each having varying levels of accuracy. Approaches for creating predictive models over large areas include linear regression, logistic regression, discriminant analysis, principal component analysis, canonical correlation analysis, genetic algorithms (e.g. GARP), classification and regression (CART) analysis, maximum entropy (MAXENT) and other Bayesian methods (Scott et al. 2002, Elith et al. 2006). There is no standard or best practice when modeling species distributions; however some analyses are better suited to particular data sets and modeling goals (Elith et al. 2006). Methods such as GARP and MAXENT may be promising modeling techniques for data that do not include species absence information (Phillips et al. 2006). If information on species absence is

available, other analyses such as regression techniques, both in standard and Bayesian frameworks, may be more useful.

Model accuracy can be compromised in many ways and may vary more between species of interest than among actual modeling techniques (Seoane et al. 2005). Error propagation is also common with multipyscaled data or any time multiple data layers are combined. This refers to the process in which errors present in spatial data are passed through a GIS operation and accumulate in output products (Veregin 1989). Larger data sources, in terms of number of data points, along with standardization of surveys, can minimize error at the start of modeling (Seoane et al. 2005). Additionally, consistency in scale across datasets will aid in preventing error propagation, although this may not properly characterize the processes of interest.

Previous modeling efforts for Cerulean Warblers have had promising results. Dettmers and Bart (1999) constructed species-specific habitat models for Cerulean Warblers and eight other forest-breeding songbirds in southern Ohio using presence-only data. Other methods of presence-only modeling have been created based on the Mahalanobis statistic (e.g., Clark et al. 1993, Knick and Dyer 1997, Buehler et al. 2006); however, the methods employed in the Ohio study were primarily GIS-based techniques that constructed quantitative descriptions of “good habitat” using habitat variables. Once optimum ranges were chosen for each variable, multivariate models were constructed using Boolean operators and parentheses. Although the spatial extent of the model was limited, model performance was better than random at identifying where each species occurred. The model was also useful for predicting the amount and spatial distribution of breeding habitat.

Wood et al (2006b) used fine-scale bird observation datasets to model and analyze presence, absence, and abundance of Cerulean Warblers and four other forest-dependent songbird species. Forty-five landscape metrics were derived from a digital elevation model (DEM) and the National Land Cover Dataset (NLCD). These metrics were calculated at three separate scales. CART models were constructed using a specified minimum of 5 observations for a node split, with a minimum deviance of 0.010 for each node, and a minimum node size of 10 observations. Overfitted trees were pruned using the sum of squared errors or squared deviation to identify breakpoints in tree size, similar to a 10-fold cross validation process. Models for all species performed moderately well with varying success. Cerulean Warbler

presence was best predicted by a presence/absence model, while an abundance model best predicted absence.

Results of CART models must be interpreted with the caveat that they are largely exploratory in nature and only return estimates of correlation. Yet trees have many advantages including the flexibility to handle a broad range of response types, ease and robustness of construction, ease of interpretation, and the ability to handle missing values in both response and explanatory variables (De'ath and Fabricius 2000). However, despite their sophistication, results are subject to the normal limitations of correlation analysis, specifically that correlation does not ensure causation (Hahn and O'Connor 2002).

Other models for Cerulean Warblers, including those by Buehler et al. (2006), Thogmartin et al. (2004), and Thogmartin (unpublished data) have had relatively high success in their geographic study areas. Each of these models identified slightly different variables as important to identifying suitable habitat for Cerulean Warblers (Table 1). Identification of different variables may be a result of differences in geographic location, time period, scale, and availability of data.

Like many other modeling efforts, the model created by Buehler et al. (2006) applied to a relatively small study area. The study area encompassed 206,579 ha of the Cumberland Mountains in four eastern Tennessee counties. Habitat variables were developed from remotely-sensed vegetation and landform data with methods similar to the modeling efforts previously mentioned. Randomly-selected transects were surveyed, and singing male Cerulean Warblers were identified and marked with a Global Positioning System unit. Models were created using the Mahalanobis distance statistic. Performance of the model with regards to Cerulean Warbler presence was high (93% correct classification of 2003 data, and 80% correct classification based on 2004 data), but absences were only correctly classified 54% of the time. This skewed classification may be a result of the survey techniques incorporated into the modeling.

Alternative modeling approaches using Bayesian inference have been employed using avian surveys and environmental variables (e.g., Link and Sauer 2002, Thogmartin et al. 2004). Although counts are the primary means by which abundance is indexed for birds, they are confounded by nuisance effects (e.g., observer, time of day, season) and spatial correlation between counts (Thogmartin et al. 2004). Most statistical models have random variables that depend on fixed parameters; however, in hierarchical models the parameters are treated as

random variables (Link and Sauer 2002). Hyperparameters determine the distribution of the parameters (Gelman et al. 2004). This structure allows for modeling the influence of nuisance effects on the distribution of the parameters influencing counts, rather than on the counts themselves (Link and Sauer 2002).

Given the observed data, Bayesian methods calculate the probability of the value of a parameter (Wade 2000). Traditional statistical analyses (frequentist methods) calculate the probability of observing the data given a specific value for a parameter (i.e., null hypothesis). Rather, Bayesian inference focuses on what the data tell about the parameter (Wade 2000).

If counts are expected to be spatially correlated, geographic neighbors may be specified *a priori*, forming a Bayesian smoothing model (Best et al. 1999). A conditional autoregressive (CAR) structure may be used in the model to model spatial similarity (Best et al. 1999, Thogmartin et al. 2004). A flat prior Bayesian model accounting for space, environmental covariates, and count structure can be implemented to identify unbiased relationships between environmental covariates and avian abundance (Link and Sauer 2002, Thogmartin et al. 2004).

Sophistication of models will ultimately depend on the management objectives, and the ability to translate those models into tools that can be used by biologists and land managers. There is a current disconnect between managers and statisticians. While biologists and modelers currently recognize that many models are archaic, the need for ease of use will always be an important factor in model construction.

This thesis will explore both classification and regression tree models and Bayesian hierarchical spatial models of Cerulean Warbler presence and abundance at multiple scales. Both techniques have been used to model abundance and presence/absence with count data (e.g., Dettmers and Bart 1999, D'earth and Fabricius 2000, Thogmartin et al. 2004). Additionally, both techniques have been used to model Cerulean Warblers and other songbirds (e.g., Link and Sauer 2002, Buehler et al. 2006, Wood et al. 2006b). The two techniques vary in complexity and ease of use, but one technique may be more suited to the structure and resolution of the data, as well as the intended management objective.

## LITERATURE CITED

Ammer, F.K. 2003. Population level dynamics of Grasshopper Sparrow populations breeding on reclaimed mountaintop mines in West Virginia. PhD Dissertation, West Virginia



- University, Morgantown. <https://eidr.wvu.edu/etd/documentdata.eTD?documentid=3242>
- Atlantic Coast Joint Venture. 2005. Appalachian BCR Concept Plan Version 1.0 July 2005.  
Atlantic Coast Joint Venture. [http://www.acjv.org/documents/bcr28\\_concept\\_plan.pdf](http://www.acjv.org/documents/bcr28_concept_plan.pdf)
- Bajema, R.A., and S.L. Lima. 2001. Landscape-level analyses of Henslow's Sparrow (*Ammodramus henslowii*) abundance in reclaimed coal mine grasslands. *American Midland Naturalist* 145:288-298.
- Baker, M.D., and M.J. Lacki. 1997. Short-term changes in bird communities in response to silvicultural prescriptions. *Forest Ecology and Management* 96:27-36.
- Beard, K.H., N. Hengartner, and D.K. Skelly. 1999. Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology* 113:1108-1116.
- Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. *American Scientist* 54: 691-692.
- Best, N. G., R. A. Arnold, A. Thomas, L. A. Waller, and E.M. Conlon. 1999. Bayesian models for spatially correlated disease and exposure data. *Bayesian Statistics* 6:131-156.
- Bibby, C.J., N.D. Burgess, and D.A. Hill. 1992. *Bird Census Techniques*. Academic Press Inc., San Diego, CA.
- Boardman, L.A., and R.H. Yahner. 1999. Wildlife communities associated with even-aged reproduction stands in two state forests of Pennsylvania. *Northern Journal of Applied Forestry* 16:89-95.
- Bonan, G.B. 1999. Frost followed the plow: impacts of deforestation on the climate of the United States. *Ecological Applications* 9:1305-1315.
- Boone, R.B., and W.B. Krohn. 2002. Modeling tools and accuracy assessment. Pages 265-270 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Bosworth, S.B. 2003. Cerulean Warbler relative abundance and frequency of occurrence relative to large-scale edge. MS Thesis, West Virginia University, Morgantown.  
[https://etd.wvu.edu/etd/etdDocumentData.jsp?jsp\\_etdId=3116](https://etd.wvu.edu/etd/etdDocumentData.jsp?jsp_etdId=3116)
- Boulinier, T., J.D. Nichols, J.E. Hines, J.R. Sauer, C.H. Flather, and K.H. Pollock. 1998. Higher temporal variability of forest breeding bird communities in fragmented landscapes. *Proceedings of the National Academy of Sciences* 95:7497-7501.
- Brittingham, M.C., and S.A. Temple. 1983. Have cowbirds caused forest songbirds to decline? *BioScience* 33:31-35.

- Buehler, D.A., M.J. Welton, and T.A. Beachy. 2006. Predicting Cerulean Warbler habitat use in the Cumberland Mountains of Tennessee. *Journal of Wildlife Management* 70:1763-1769.
- Bulluck, L.P., and D.A. Buehler. 2006. Avian use of early successional habitats: are regenerating forests, utility right-of-ways and reclaimed surface mines the same? *Forest Ecology and Management* 236:76-84.
- Clark, J.D., J.E. Dunn, and K.G. Smith. 1993. A multivariate model of female black bear habitat use for a geographic information system. *The Journal of Wildlife Management* 57:519-526.
- COSEWIC. 2003. COSEWIC assessment and update status report on the Cerulean Warbler *Dendroica cerulea* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa.vii + 25 pp.
- De'ath, G., K.A. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.
- Dettmers, R., and J. Bart. 1999. A GIS modeling method applied to predicting forest songbird habitat. *Ecological Applications* 9:152-163.
- Duguay, J.P., P.B. Wood, and J.V. Nichols. 2001. Songbird abundance and avian nest survival rates in forests fragmented by different silvicultural treatments. *Conservation Biology* 15:1405-1415.
- Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods for predicting species distribution from occurrence data. *Ecography* 29:129-151.
- Fearer, T.M., S.P. Prisley, D.F. Stauffer, and P.D. Keyser. 2007. A method for integrating the breeding bird survey and forest inventory and analysis databases to evaluate forest bird-habitat relationships at multiple spatial scales. *Forest Ecology and Management* 243:128-143.
- Federal Register. 2006. Endangered and threatened wildlife and plants; 12-moth finding on a petition to list the Cerulean Warbler (*Dendroica cerulea*) as threatened with critical habitat. 71(234):70717-70733.
- Fitzgerald, J. A., W. E. Thogmartin, R. Dettmers, T. Jones, C. Rustay, J. M. Ruth, T. C. Will, and F. R. Thompson III. 2008. Models for conservation planning for terrestrial birds in North America. Pages 593–624 in J. J. Millspaugh and F. R. Thompson III, editors.

- Models for planning wildlife conservation in large landscapes. Academic/Elsevier, Boston, Massachusetts, USA/Amersterdam, Netherlands.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2004. Bayesian Data Analysis. Chapman and Hall, New York, New York, USA.
- Gutzwiller, K.J., and W.C. Barrow Jr. 2001. Bird-landscape relations in the Chihuahuan Desert: coping with uncertainties about predictive models. *Ecological Applications* 11:1517-1532.
- Hahn, D.C., and R.J. O'Connor. 2002. Contrasting determinants of abundance in ancestral and colonized ranges of an invasive brood parasite. Pages 219-228 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Hamel, P.B. 2000. Cerulean Warbler (*Dendroica cerulea*). In *The Birds of North America*, No. 511 (A. Poole and F. Gill, eds.). The Birds of North America, Inc., Philadelphia, PA.
- Hamel, P.B., D.K. Dawson, and P.D. Keyser. 2004. How we can learn more about the Cerulean Warbler (*Dendroica cerulea*). *Auk* 121:7-14.
- Hartman, P.J. 2006. Habitat selection of the Cerulean Warbler in Eastern Kentucky. MS Thesis, University of Kentucky, Lexington.  
<http://archive.uky.edu/bitstream/10225/306/HartmanThesis.pdf>
- Heglund, P.J. 2002. Foundations of species-environment relations. Pages 35-41 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Hunter, W.C., D.A. Buehler, R.A. Canterbury, J.L. Confer, and P.B. Hamel. 2001. Conservation of disturbance-dependent birds in eastern North America. *Wildlife Society Bulletin* 29:440-455.
- Hutson, M.A. 2002. Introductory Essay: Critical issues for improving predictions. Pages 7-21 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Johnson, B.R., W.F. Cross, and J.B. Wallace. 2003. Long-term resource limitation reduces insect detritivore growth in a headwater stream. *Journal of the North American Benthological Society* 22:565-574.
- Jones, J., R.D. DeBruyn, J.J. Barg, and R.J. Robertson. 2001. Assessing the effects of natural disturbance on a Neotropical migrant songbird. *Ecology* 82:2628-2635.
- Keller, C.M.E., J.T. Scallan. 1999. Potential roadside biases due to habitat changes along breeding bird survey routes. *Condor* 101:50-57.

- King, D.I., and DeGraaf, R.M. 2000. Bird species diversity and nesting success in mature, clearcut and shelterwood forest in northern New Hampshire, USA. *Forest Ecology and Management* 129:227-235.
- Knick, S.T., and D.L. Dyer. 1997. Distribution of black-tailed jackrabbit habitat determined by GIS in Southwestern Idaho. *Journal of Wildlife Management* 61:75-85.
- LaDeau, S.L., A.M. Kilpatrick, and P.P. Marra. 2007. West Nile virus emergence and large-scale declines of North American bird populations. *Nature* 447:710-714.
- Levin, S.A. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943-1967.
- Li, H., and J. Wu. 2004. Use and misuse of landscape indices. *Landscape Ecology* 19:389-399.
- Link, W.A., and J.R. Sauer. 2002. A hierarchical analysis of population change with application to Cerulean Warblers. *Ecology* 83:2832-2840.
- Litvaitis, J.A. 1993. Response of early successional vertebrates to historic changes in land use. *Conservation Biology* 7:866-873.
- Loveland, T.R., and W. Avcedo. 2006. Land cover change in the Eastern United States. in *Status and Trends of Eastern United States Land Cover*. United States Geological Survey. <http://edc2.usgs.gov/LT/LCCEUS.php>
- Lowther, P.E. 1993. Brown-headed Cowbird (*Molothrus ater*). In *The Birds of North America*, No. 47 (A. Poole and F. Gill Eds.). Philadelphia: The Academy of Natural Sciences; Washington, D.C.: The American Ornithologists' Union.
- Lutz, W., and R. Qiang. 2002. Determinants of human population growth. *Philosophical Transactions: Biological Sciences* 357(1425):1197-1210.
- Martel, J., and Y. Mauffette. 1997. Lepidopteran communities in temperate deciduous forests affected by forest declines. *Oikos* 78:48-56.
- Maurer, B.A., L.B. McArthur, and R.C. Whitmore. 1981. Effects of logging on guild structure of a forest bird community in West Virginia. *American Birds* 35:11-13.
- Maurer, B.A. 1996. Relating human population growth to the loss of biodiversity. *Biodiversity Letters* 3:1-5
- McDermott, M. E., and P. B. Wood. 2009. Short- and long-term implications of clearcut and two-age silviculture for conservation of breeding forest birds in the central Appalachians, USA. *Biological Conservation* 142:212-220.
- McPherson, J.M., W. Jetz, and D.J. Rogers. 2006. Using coarse-grained occurrence data to

- predict species distributions at finer spatial resolutions – possibilities and limitations. *Ecological Modelling* 192:499-522.
- Miller, G. W., J. N. Kochenderfer, and D. B. Fekedulegn. 2006. Influence of individual reserve trees on nearby reproduction in two-aged Appalachian hardwood stands. *Forest Ecology and Management* 224:241-251.
- Muller, K.L., J.A. Stamps, V.V. Krishnan, and N.H. Willits. 1997. The effects of conspecific attraction and habitat quality on habitat selection in territorial birds (*Troglodytes aedon*). *The American Naturalist* 150:650-661.
- O’Conner, R.J., M.T. Jones, D. White, C. Hunsaker, T. Loveland, B. Jones, and E. Preston. 1996. Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States. *Biodiversity Letters* 3:97-110.
- Perkins, K.A. 2006. Cerulean Warbler selection of forest canopy gaps. MS Thesis, West Virginia University, Morgantown. <https://etd.wvu.edu/etd/documentdata.eTD?documentid=4596>
- Peterjohn, B.G. 2001. Some considerations on the use of ecological models to predict species’ distributions. *Condor* 103:661-663.
- Phillips, S.J., R.P. Anderson, R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.
- Ramankutty, N., and J.A. Foley. 1999. Estimating historical changes in land cover: North American croplands from 1850 to 1992. *Global Ecology and Biogeography* 8:381-396.
- Riitters, K.H., R.V. O’Neill, and K.B. Jones. 1997. Assessing habitat suitability at multiple scales: a landscape-level approach. *Biological Conservation* 81:191-202.
- Robbins, C.S., J.W. Fitzpatrick, and P.B. Hamel. 1992. A warbler in trouble: *Dendroica cerulea*. Pages 549–562 in *Ecology and Conservation of Neotropical Migrant Landbirds* (J. M. Hagan III and D. W. Johnston, Eds.). Smithsonian Institution Press, Washington, D.C.
- Robinson, S.K., F.R. Robinson III, T.M. Donovan, D.R. Whitehead, and J. Faaborg. 1995. Regional forest fragmentation and the nesting success of migratory birds. *Science* 267:1987-1990.
- Rosenberg, K.V., S.E. Barker, and R.W. Rohrbaugh. 2000. An atlas of Cerulean Warbler populations. Final Report to the U.S. Fish and Wildlife Service, December 2000.
- Sauer, J. R., J. E. Hines, and J. Fallon. 2008. The North American Breeding Bird Survey, Results and Analysis 1966 - 2006. Version 10.13.2007. USGS Patuxent Wildlife Research Center, Laurel, MD <http://www.mbr-pwrc.usgs.gov/bbs/bbs.html>
- Sauer, J.R., B.G. Peterjohn, and W.A. Link. 1994. Observer differences in the North American

Breeding Bird Survey. *Auk* 111:50-62.

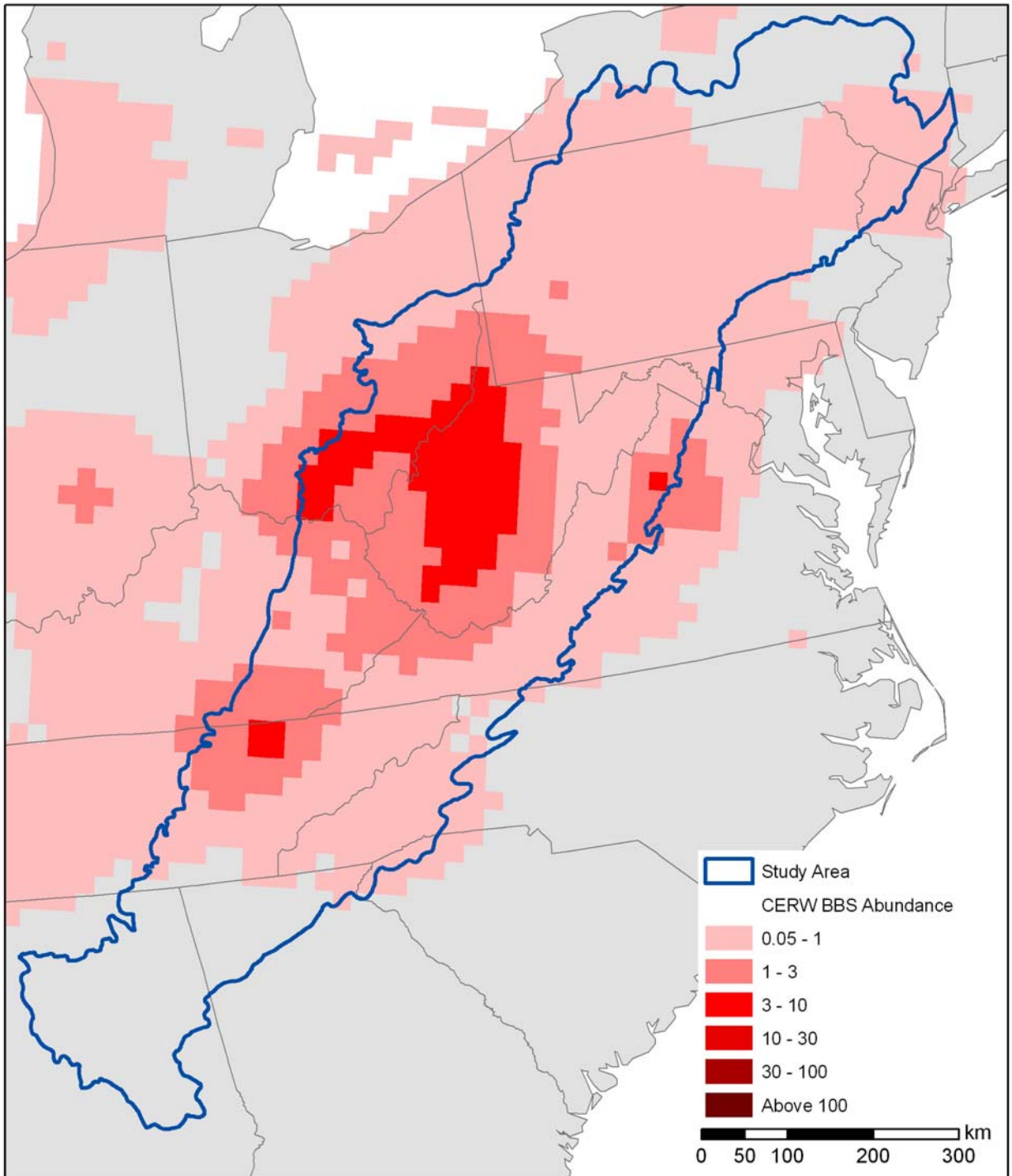
- Schaefer, S.M., and W.B. Krohn. 2002. Predicting vertebrate occurrences from species habitat associations: Improving the interpretation of commission error rates. Pages 419-427 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Scott, J.M., J.H. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson, editors. 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Seoane, J., L.M. Carrascal, C.L. Alonso, and D. Palomino. 2005. Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecological Modelling* 185:299-308.
- Smith, H. C., N. I. Lamson, and G. W. Miller. 1989. An esthetic alternative to clearcutting? *Journal of Forestry* 87:14-18.
- Stein, S.M., R.E. McRoberts, R.J. Alig, M.D. Nelson, D.M. Theobald, M. Eley, M. Dechter, and M. Carr. 2005. *Forests on the edge: Housing development on America's private forests*. Portland, OR. General Technical Report PNW-GTR-636. United States Department of Agriculture, 16 p. <http://www.fs.fed.us/projects/fote/reports/fote-6-9-05.pdf>:
- Thogmartin, W.E., J.R. Sauer, and M.G. Knutson. 2004. A hierarchical spatial model of avian abundance with application to Cerulean Warblers. *Ecological Applications* 14:1766-1779.
- Thogmartin, W.E., F.P. Howe, F.C. James, D.H. Johnson, E.T. Reed, J.R. Sauer, and F.R. Thompson III. 2006. A review of the population estimation approach of the North American Landbird Conservation Plan. *Auk* 123:892-904.
- Thogmartin, W.E., J.A. Fitzgerald, C.A. Drew, and M.T. Jones. 2008. Conservation design: Where do we go from here? Proceedings of the 4th International Partners in Flight Conference, McAllen, Texas, USA.
- Tobalske, C. 2002. Effects of spatial scale on the predictive ability of habitat models for the Green Woodpecker in Switzerland. Pages 197-204 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Turner, M.G., V.H. Dale, and R.H. Gardner. 1989. Predicting across scales: Theory development and testing. *Landscape Ecology* 3:245-252.
- US EPA. 2005. Mountaintop mining/valley fills in Appalachia: final programmatic environmental impact statement (FPEIS). Philadelphia, PA Report no. EPA 9-03-R-05002. <http://www.epa.gov/region3/mtntop/eis.htm>

- Veregin, H. 1989. Error modeling for the map overlay operation. In M. Goodchild & S. Gopal (Eds.), *The Accuracy of Spatial Databases* (pp. 3-18). London: Taylor & Francis.
- Wade, P.R. 2000. Bayesian methods in conservation biology. *Conservation Biology* 14:1308-1316.
- Weakland, C.A., and P.B. Wood. 2005. Cerulean Warbler (*Dendroica cerulea*) microhabitat and landscape-level habitat characteristics in southern West Virginia. *Auk* 122:497-508.
- West Virginia Division of Highways. 2003. The Route – Corridor H.  
<http://www.wvcorridorh.com/route/route.html>
- Wickham, J.D., K.H. Riitters, T.G. Wade, M. Coan, and C. Homer. 2006. The effect of Appalachian mountaintop mining on interior forest. *Landscape Ecology* 22:179-187.
- Will, T.C., J.M. Ruth, K.V. Rosenberg, D. Krueper, D. Hahn, J. Fitzgerald, R. Dettmers, and C.J. Beardmore. 2005. The five elements process: Designing optimal landscapes to meet bird conservation objectives. Partners in Flight Technical Series No. 1.  
<http://www.partnersinflight.org/pubs/ts/01-FiveElements.pdf>
- Wood, P.B., J.W. Edwards, and C.A. Weakland. 2001. Terrestrial vertebrate (breeding songbird, raptor, small mammal, herpetofaunal) populations of forested and reclaimed sites. Final report submitted to EIS Steering Committee.  
<http://www.epa.gov/region03/mtntop/pdf/Appendices/Appendix%20E%20Terrestrial/Vertebrate%20study/vertebratestudy.pdf>
- Wood, P.B., S.B. Bosworth, and R. Dettmers. 2006a. Cerulean Warbler abundance relative to large-scale edge and habitat characteristics. *Condor* 108:154-165.
- Wood, P.B., M.P. Strager, J.M. Strager. 2006b. Fine-scale forest bird habitat modeling for the mountaintop mining region within the Appalachian Bird Conservation Region. Final Report to the USGS, August 2006.
- Zabel, C.J., L.M. Roberts, B.S. Mulder, H.B. Stauffer, J.R. Dunk, K. Wolcott, D. Solis, Mike Gertsch, B. Woodbridge, A. Wright, G. Goldsmith, and C. Keckler. 2002. A collaborative approach in adaptive management at a large-landscape scale. Pages 241-253 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.

**Table 1.** Key variables from previous Cerulean Warbler modeling efforts.

<b>Variable</b>	<b>Dettmers and Bart 1999</b>	<b>Thogmartin et al. 2004</b>	<b>Buehler et al. 2006</b>	<b>Wood et al. 2006b</b>	<b>Thogmartin unpublished data</b>
elevation			x	x	
slope	x		x	x	
slope position	x		x	x	
surface morphology	x				
solar exposure			x	x	
moisture index	x	x			
water flow accumulation	x				
distance to stream			x	x	
developed landcover				x	
forested landcover	x		x		x
core forest / ↑ forest patch size		x		x	
forest edge density					x
percent forested wetland		x			
deciduous forest patch size		x			
soil pH / acid deposition					x
precipitation		x			x





**Figure 1.** Breeding range of Cerulean Warbler as derived from an interpolation of Breeding Bird Survey counts (Sauer et al. 2007).

## **Chapter 2**

### **Comparison of classification and regression trees and Bayesian modeling approaches for predicting the occurrence of cerulean warblers**

**Formatted in the style of *The Journal of Wildlife Management***

## **ABSTRACT**

Many statistical approaches have been used for developing predictive models for wildlife presence/absence and abundance, each with varying levels of accuracy and complexity. As concerns for declining species intensify and anthropogenic impacts on habitats increase, the ability to quickly quantify and map species distributions and abundances over large regions will become increasingly important. To date, there is no set of best practices for modeling specific wildlife groups. Here I present a comparison of two modeling techniques that range from moderately quick and simplistic (decision trees) to conceptually and computationally complex (hierarchical spatial models). I used North American Breeding Bird Survey counts with a suite of explanatory variables to predict presence and abundance of cerulean warblers (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region. Of the decision tree methods, cerulean warbler occurrence was most accurately described by presence/absence models. Regression tree abundance models under-predicted counts and had low accuracy. Hierarchical spatial models predicted abundance of cerulean warblers similar to actual counts, and with better overall accuracy than regression trees. All techniques produced models using similar variables; interior forest and percent forest were most important for identifying areas with cerulean warblers. The results and relationships identified provide a useful comparison of these modeling techniques and may serve to direct management and monitoring of areas of conservation significance for cerulean warblers.

**KEYWORDS** Bayesian, cerulean warbler, classification and regression trees, count data, *Dendroica cerulea*, Markov chain Monte Carlo, North American Breeding Bird Survey.

## INTRODUCTION

Biologists and land managers must often prioritize which specific lands and habitat types need to be restored, protected, and/or managed to most effectively achieve conservation objectives for species of interest (Will et al. 2005, Fitzgerald et al. 2008, Thogmartin et al. 2008). A fundamental step in this process is the ability to accurately predict species occurrence on the landscape. This is an important task, particularly as less habitat may be available, and managing for declining species necessitates optimization of land use and limited conservation resources. As our technology advances and as range-wide species survey data and remotely sensed environmental data become more available, the ability to accurately and efficiently model systems will become more important. However, the choice between accuracy and simplicity is a tradeoff in many modeling processes. Though attempts at describing ecosystems and mapping species across the landscape are well documented (Scott et al. 2002, Guisan and Thuiller 2005, Elith et al. 2006), we continue to struggle with the concept and application. One of the major problems plaguing species models is the lack of basic, yet crucial, life history information and population demographics. To understand how organisms' populations and distributions vary over time and space, we need to find appropriate ways to quantify patterns (Wiens 1989).

Vertebrates, particularly birds, have been well represented in past modeling efforts (Verner et al. 1986). Yet, to date many avian models have had relatively poor predictability rates because of problems with differences in species' habitat occurrence, issues of scale, sampling biases, lack of knowledge concerning species' life histories, and many other factors (Beard et al. 1999, Gutzwiller and Barrow 2001, Wood et al. 2006). Aside from these issues, many modeling methods poorly accommodate the nature of counts, observer and spatial effects, and other effects associated with survey design and sampling schemes (Thogmartin et al. 2004). These biases can present overly optimistic or pessimistic estimates (Bibby et al. 1992, Sauer et al. 1994, Keller and Scallan 1999, Peterjohn 2001). Furthermore, there is an underlying assumption that observations equate to patterns (Li and Wu 2004), yet in some cases, this may be an artifact of the scale of analysis (McPherson et al. 2006). Changes in scale may result in a change in observed patterns (Levin 1992, Riitters et al. 1997, McPherson et al. 2006). Patterns observed at a particular scale may be the result of processes operating at another scale or the combination of a suite of scales (McPherson et al. 2006, Thogmartin 2007).

There are many statistical approaches for model creation, each having varying levels of accuracy, yet there is no standard or best practice when modeling species distributions. However, some analyses are better suited to particular data sets and modeling goals (Elith et al. 2006). Methods such as MAXENT may be a promising modeling technique for data that do not include species absence information (Phillips et al. 2006). If information on species absence is available, regression techniques, both in standard and Bayesian frameworks, may be more useful (Elith et al. 2006).

Classification and regression trees (CART) have successfully been used to predict habitat use and distributions for a variety of taxa including songbirds (O’Conner et al. 1996, Wood et al. 2006), mammals (O’Brien et al. 2005), and plants (Bourg et al. 2005). CART, a data-mining technique, operates by recursive partitioning thereby creating decision trees that are similar in concept to dichotomous trees (Vayssières et al. 2000). Terminal nodes of the tree result in a classification of the response variable. Decision trees have performed as well as other techniques such as logistic regression, ANOVA, and linear regression, yet may be better at revealing patterns in datasets (Vayssières et al. 2000, De’ath and Fabricius 2000, Bourg et al. 2005). Results of CART models must be interpreted with the caveat that they are largely exploratory in nature and only return estimates of correlation. Yet trees have many advantages including the flexibility to handle a broad range of response types, ease and robustness of construction, ease of interpretation, and the ability to handle missing values in both response and explanatory variables (De’ath and Fabricius 2000). However, despite their sophistication, results are subject to the normal limitations of correlation analysis, specifically that correlation does not ensure causation (Hahn and O’Connor 2002).

Alternative modeling approaches using Bayesian inference have been based on avian surveys and environmental variables (e.g., Link and Sauer 2002, Thogmartin et al. 2004). Although counts are the primary means by which abundance is indexed for birds, they are confounded by nuisance effects (e.g., observer, time of day, season) and spatial correlation between counts (Thogmartin et al. 2004). Most statistical models have random variables that depend on fixed parameters; however, in hierarchical models the parameters are treated as random variables (Link and Sauer 2002). Hyperparameters determine the distribution of the parameters (Gelman et al. 2004). This structure allows for modeling the influence of nuisance

effects on the distribution of the parameters influencing counts, rather than on the counts themselves (Link and Sauer 2002).

Thogmartin et al. (2004) used a Bayesian technique to create a predictive model of abundance for the cerulean warbler in the Prairie Hardwood Transition Bird Conservation Region (BCR23) of the upper Midwestern United States. The hierarchical spatial model accounted for habitat, spatial relatedness of count locations, year effects, and observer effects. Conceptually and computationally more complex than decision tree analyses, the model was able to overcome problems with extra-Poisson dispersion of counts, nuisance effects associated with count data collection, and spatial autocorrelation.

Sophistication of models will ultimately depend on management objectives, and the ability to translate those models into tools that can be used by biologists and land managers. Here I present a comparison of two modeling techniques for presence/absence and relative abundance data that range from moderately simplistic (classification and regression trees) to conceptually and computationally complex (hierarchical spatial models). To apply the models, I predicted presence/absence and relative abundance for the cerulean warbler (*Dendroica cerulea*) from North American Breeding Bird Survey (NABBS) data for the Appalachian Mountains Bird Conservation Region (BCR28) in the eastern United States. The cerulean warbler, a small (11.5 cm, 8-10 g) Neotropical migrant that breeds in mature deciduous forest, has its core range in the central Appalachians (Figure 1). A rangewide population decline has made it a focal species of research and management priority (Robbins et al. 1992, Hamel et al. 2004). My specific objectives were to determine 1) the predictive accuracy of each type of model and 2) the ease of implementing each technique.

## **STUDY AREA**

The Appalachian Bird Conservation Region (BCR28) comprises >426,000 km<sup>2</sup> throughout 15 states. The portion that overlaps Massachusetts and Connecticut was removed prior to analyses (Figure 2); because of its small width to length ratio, it does not completely contain any North American Breeding Bird Survey (NABBS) routes. The study region intersects or contains 15 Partners in Flight (PIF) physiographic regions (Figure 2).

The study region is dominated by oak-hickory (*Quercus* and *Carya* spp.), mixed-mesophytic, and northern hardwood forest types (ACJV 2005). Topography consists of mostly

mountainous regions throughout the Appalachian Mountains. Other types of habitat include high elevation red spruce (*Picea rubens*) forests, wetlands, scrublands, and grasslands. The diverse habitat types are partly a result of the large range in latitude, longitude, and elevation (0-1,981 m above sea level).

## METHODS

### *Data sets*

Models were constructed using NABBS data. The NABBS is a long-term roadside survey initiated in 1966 to monitor bird population trends within North America (Sauer et al. 2008). Surveys are completed annually during May and June along randomly assigned roadside routes. Each route contains 50 stops spaced approximately 0.8 km along the route. Observers conduct a 3 minute point count at each stop and record all birds seen or heard within a 402 m radius. Surveys begin 30 minutes before local sunrise and take approximately 5 hours to complete. When possible, individual routes are surveyed by the same observer each year and under suitable weather conditions (i.e., low wind, low-precipitation, high visibility). The protocol is designed to maintain a level of consistency and minimize sampling variability so that variations in population trends can be detected over time.

I used count data from all NABBS routes within BCR28 that were active between 1998 and 2004 ( $n = 348$ ) to develop and validate models. The years 1998-2004 comprise a 7-year window that coincides with the time period of the 2001 National Landcover Dataset (NLCD, Homer et al. 2004). For hierarchical spatial analyses, abundance was the sum of cerulean warbler counts over the 50 stops each year. When fewer than 50 stops were surveyed on a route in a given year, the sum was increased in proportion to the number of stops surveyed. For example, if a route had 48 stops with a total of 10 cerulean warblers observed, the inflated sum for the route would equal 10.42 ( $10 + (10/48)*(50-48)$ ). The sum was averaged over the 7-year period for decision tree analyses. I randomly selected 30% of the data to be withheld for validation.

I reviewed previous modeling efforts for cerulean warblers (Dettmers et al. 1999, Thogmartin et al. 2004, Buehler et al. 2006, Wood et al. 2006) and relied on expert opinion to select a parsimonious suite of environmental covariates *a priori* that seemed biologically significant to the focal species (Table 1). I derived these environmental covariates from

moderate-resolution remotely-sensed data with a grain of 30 m x 30 m and primarily represented land cover. Most topographic variables (i.e., slope, aspect, slope position, moisture) were not included because they were not appropriate for use over large geographic areas. Topographic variables such as aspect will encompass most values over a large geographic area, thus will average out to some uninformative orientation. Additionally, I did not include climatic variables for either set of models to focus on variables that could more easily be implemented in management activities.

Metrics were calculated within each of three buffers (100 m, 300 m, and 1000 m) around each NABBS route using ArcGIS® version 9.1 (ESRI 2006) and the Spatial Analyst (ESRI 2006) and Patch Analyst© version 4 (Ontario Ministry of Natural Resources 2008) extensions. Buffer sizes were chosen to represent 3 unique scales relevant to cerulean warbler biology (specifically, extent or level of analysis); immediate resource area, local, and regional, respectively. The smallest extent is comparable to the nest site and immediate vicinity, the habitat most used by a territorial adult, while local or intermediate would represent the maximum territory size, and regional extent is an area over which the species may select breeding habitat.

I used the USGS National Elevation Dataset (NED) for elevation. Land cover metrics were derived from the National Landcover Dataset (NLCD, Homer et al. 2004). Distance to forested streams was calculated from a combination of NLCD and National Hydrography Dataset (NHDPlus: 1:100,000 scale). Because I selected a parsimonious suite of environmental covariates in the models and the data were restricted to a 30 m cell size resolution, some variables important to cerulean warblers may have been omitted. To help accommodate this, I included latitude and longitude in the modeling as a surrogate. This also may help to address some of the geographic patterning in the dataset (Franklin 1998).

### *Decision tree models*

I used CART to model presence/absence and relative abundance for cerulean warblers. Decision trees were created from the data set for two separate analyses: (1) classification tree analysis of presence/absence and (2) regression tree analysis of abundance. Cerulean warbler presence/absence and abundance were modeled as a function of environmental covariates. Nuisance variables (e.g., observer and route) could not be included in the models. Analyses were performed in R (R 2008) using the MVMART and CARET packages. I used multiple (50) 10-fold cross-validation to create “pruned” or optimal tree sizes. During each of the 50 runs, the



data are divided into 10 subsamples. Each subsample is used once for training and validation purposes (Vayssieres et al. 2000, Bourg et al. 2005). Overgrown trees can reduce the misclassification error of the data set used for construction, but perform poorly on independent validation data. In addition to the multiple cross-validation, I chose final tree sizes that were within 1 standard error of the minimum cross-validation (Breimen et al. 1984). Data from the 104 routes withheld from model construction were used to validate model performance.

### *Hierarchical spatial models*

I modeled relative bird abundance as a function of environmental covariates in hierarchical spatial count models as per Thogmartin et al. (2004). I implemented a flat prior (no prior information) Bayesian model that accounts for count structure, environmental covariates, and space (Gilks et al. 1996, Spiegelhalter et al. 2000, Link and Sauer 2002, Gelmen et al 2004, Thogmartin et al. 2004). I used a Poisson model because count data typically have a Poisson distribution (Cunningham and Lindenmayer 2005). The spatial Poisson model was specified as

$$\log[\lambda(s)] = \sum_{k=1}^p \beta_k \chi_k(s) + Z_k(s) + \omega_k(s) + \gamma_k(s) + \varepsilon_k$$

where  $x(s)$  are spatially indexed environmental covariates and  $\beta_k$  is the change in abundance (on the log scale) per unit change in covariate  $k$  (Royle et al. 2002). The index  $s$  is in geographic coordinates (Thogmartin et al. 2004). The random effects  $Z(s)$ ,  $\omega(s)$  and  $\gamma(s)$  control for the spatial relatedness of counts and nuisance effects for observer and year, respectively. All environmental covariates were treated as fixed effects and assumed no error in measurement (Thogmartin et al. 2004). Extra-Poisson dispersion was accounted for by the overdispersion effect  $\varepsilon_k$ .

To determine the spatial relatedness of counts, I first created a spatial neighborhood for NABBS routes. I delineated the spatial neighborhood by tessellating the surveys and created an adjacency matrix on the derived irregular lattice (Figure 3) using the Fox Adjacency Tool (W.E. Thogmartin and T.J. Fox, USGS, unpublished report). Tessellation of the surveys and the application of the Fox Adjacency Tool were applied in ArcGIS® version 9.1 (ESRI 2006). Neighborhood weights were set to 1 for areas that shared a boundary and 0 for all other areas (Besag et al. 1991, Thogmartin et al. 2004). The spatial model is thus spatially dependent on the neighborhood structure and not on a distance metric. A conditional autoregressive (CAR) prior

distribution was applied to the spatial neighborhood of surveys to account for the spatial effects [Z(s)]. For additional details on the spatial model see Thogmartin et al. (2004).

I fit models using Markov Chain Monte Carlo (MCMC) techniques using Gibbs sampling (Link et al. 2002, Gelmen et al. 2004, Thogmartin et al. 2004). Three MCMC chains were computed for each simulation and tested for convergence using the Gelmen Rubin test (Brooks and Gelman 1998). Dependence on the prior is minimized during convergence which allows for the data to primarily influence the posterior distribution (Brooks and Gelman 1998). MCMC simulations were run for 100,000 simulations, including a 10,000 iteration burn in period (per chain). All variables were standardized to have zero mean and unit variance to improve MCMC performance (Gilks and Roberts 1996) and to assess comparative effect of each covariate from the model (Thogmartin et al. 2004).

Individual models were constructed at each of the three spatial scales. I used the deviance information criterion (DIC; a Bayesian measure of fit for MCMC sampling) to rank models relative to each other and averaged the environmental variables across spatial scales (Burnham and Anderson 2002, Spiegelhalter et al. 2002).

#### *Model evaluation*

I withheld a randomly chosen 30 percent of the data to validate the final models. This amounted to 104 observations (routes averaged over all years) for CART analyses and 594 observations (based on yearly data for 104 routes) for hierarchical spatial models.

For classification trees, I calculated the misclassification rate based on the original model development points, confusion matrices, and Cohen's kappa statistic. I also calculated an overall error rate based on the reserved validation data points. The kappa statistic is a measure of the correspondence between model predictions and actual observations and is more stringent than simply using the overall error rate (Landis and Koch 1977, Congalton 1991). For regression trees, I evaluated model performance using simple linear regression to test the predicted set of values against the actual observations.

Results from the hierarchical spatial models were evaluated similarly to the regression tree models. Estimated relative abundances for withheld routes from the best model were compared with known abundances using linear regression (Thogmartin et al. 2004). Regression analyses were performed using R (R 2008).

### *Mapping*

I used the final models to create spatial maps of cerulean warbler presence/absence and relative abundance in BCR28 for each modeling technique. Results from decision trees were incorporated into Spatial Analyst (ESRI 2006) as a set of nested conditional statements. Data layers were first resampled to represent the scales at which each variable was measured. Grids were resampled using neighborhood statistics with a moving window corresponding to each scale (Wood et al. 2006). For the hierarchical spatial models, data layers were standardized to match the data required for the MCMC simulations. Using the resampled grids created for the decision tree mapping, each cell was calculated using the same standardization ([value – overall mean] / standard deviation) as for the MCMC simulations (Thogmartin et al. 2007). A model averaging approach (Burnham and Anderson 2002) was then used to combine final models at individual scales to create a final map of predicted abundance for cerulean warblers throughout the Appalachian Bird Conservation Region (see Shumar chapter 3 for detailed methods).

## **RESULTS**

### *Decision trees*

Of the two CART models, cerulean warbler occurrence was most accurately predicted with models of presence/absence. The final classification tree for presence/absence had 5 splits and 6 terminal nodes, with both land cover and topographic variables represented (Figure 4). All three scales of analysis were represented in the decision tree, as well as latitude. Cross-validation with test data resulted in a model with an overall classification accuracy of 78% (Table 2) and a moderate kappa score (0.53). Classification accuracy of presence points was 73%, while accuracy for absence points was 81%. Cerulean warblers were predicted to occur throughout much of the BCR, with the highest concentrations in West Virginia, Kentucky and Ohio (Figure 5).

The final regression tree for abundance had only 1 split and 2 terminal nodes (Figure 6). Interior forest at the regional scale was the only variable used in discriminating the data. Areas with highest abundance primarily occurred in West Virginia and surrounding states (Figure 7). Data withheld from model construction suggested low correspondence between actual and estimated counts ( $R^2 = 0.16$ ; Figure 8).

### *Hierarchical spatial models*

I used a model averaging approach to combine final models at each scale (Table 3) to create a map of predicted abundance for cerulean warblers (Figure 9). The model for the 1000 m buffer scale had the highest model weight,  $w_i = 0.539$ . Models for the 300 m and 100 m buffer scales had lower weights ( $w_i = 0.330$  and  $0.130$ , respectively; Table 4). Cerulean warbler abundance increased with increasing percent forest at all scales, and declined with increasing elevation at all scales. The trend term ( $\beta_0 = -0.027$ ) was negative, indicating a decrease in cerulean warbler abundance during the time period. The route random effect explained the most variation in the data (Table 5).

Data withheld from model construction suggested better correspondence between actual and estimated counts for hierarchical spatial models ( $R^2 = 0.71$ ; Figure 8) than the regression tree model ( $R^2 = 0.16$ ; Figure 8). While there were no direct comparisons of classification trees to regression trees and hierarchical spatial models in terms of accuracy, predictive variables in final models were similar.

## **DISCUSSION**

Previous modeling efforts for cerulean warblers have had promising results, but they were restricted to small study areas (e.g., Dettmers et al. 1999, Buehler et al. 2006). Wood et al. (2006) used fine-scale bird observation datasets to model and analyze presence/absence and abundance of cerulean warblers and four other forest-dependent songbird species. CART models were constructed, and overfitted trees were pruned using the sum of squared errors or squared deviation to identify breakpoints in tree size. Models for all species performed moderately well with varying success. Cerulean warbler presence was best predicted by a presence/absence model, while an abundance model best predicted absence. Models from my study showed similar results. Overall accuracy was higher for presence/absence classification trees and hierarchical spatial models than regression tree models. Unlike most studies, Thogmartin's (W.E. Thogmartin, USGS, unpublished data) was not restricted to a small geographic area. He found that areas with the highest concentrations of cerulean warblers occurred in West Virginia, which agrees with the map created from my modeling results.

The appropriate modeling technique for a given situation will vary depending on the desired information and ultimate management objective. Classification trees for

presence/absence have high accuracy, but regression trees (created with relative abundance data) were unable to produce a model that performed well with validation data. Cerulean warbler abundance was highly kurtotic on NABBS routes throughout the Appalachian Bird Conservation Region. Though CART analyses make no assumptions about the variable distributions or data structure (De'ath and Fabricius 2000), the distribution of the data may be an important factor that is preventing the regression tree analysis from producing a model that validates well with external data.

The hierarchical spatial models performed better with the validation data than the regression tree models. Regression trees predicted lower counts on average (mean = 0.51) than were actually observed (mean = 1.32) with little correspondence to actual data, while the hierarchical spatial models predicted counts of cerulean warblers (mean = 0.68) similar to actual counts (mean = 0.67) and with better overall accuracy.

All techniques produced models identifying similar variables for cerulean warbler habitat use. Interior forest and percent forest in the landscape were important in most models for identifying areas with cerulean warblers. Other smaller-scale studies have found amount of forest in the landscape (Bosworth 2003, Buehler et al. 2006) to be a key variable for identifying areas with Cerulean Warblers. Both variables could be useful for habitat conservation and restoration. Important management questions that agencies constantly must deal with include: What habitats should be protected, and how much habitat is necessary to adequately support the species of interest? The modeling techniques discussed here highlight two approaches to answer these questions.

Classification trees of presence/absence proved to be an accurate method for identifying where cerulean warblers occurred on the landscape. Both Type I errors, where cerulean warblers were absent but predicted to occur, and Type II errors, where cerulean warblers were present but not predicted to occur, were minimal (Table 2). This technique is relatively quick and simple to apply and may be an important first step in identifying areas of concern, particularly if predicted maps are overlaid with maps of land stewardship. It should be noted however, that this technique may not provide enough information to prioritize one set of lands over another. Identifying use alone does not have the ability to identify areas of high abundance for a particular species that may be crucial to management plans. The hierarchical spatial models analyzed in this report may be a more useful technique when attempting to identify areas with high

concentrations of a species. The hierarchical spatial models overall predicted higher abundances than regression trees for cerulean warblers, especially in areas known to represent the core breeding range (Figure 10). This may be resulting from the hierarchical spatial models using an estimate for each variable in the map calculation, while only variables used in tree partitioning are used in the map creation for regression tree models. Regression trees generally predicted higher counts for cerulean warblers in other areas throughout the BCR, which may not be representative of abundance throughout the region.

Classification trees provided a fast and accurate method for identifying areas of habitat use. However, the addition of a finer resolution of response data (abundance) greatly decreased the accuracy of the models. The highly kurtotic nature of avian count data may be preventing models of abundance from fitting the data well and from performing well on validation data sets. Poisson trees may be a more useful method for datasets with the majority of counts represented by absences (Choi et al 2005). More research should be directed at improving the algorithms for differing data distributions.

Hierarchical spatial models have many benefits over a more simplistic technique such as decision tree analysis. The ability to model trends in annual counts and random effects are advantages. In my study, the trend term was negative for all models, indicating that cerulean warblers were decreasing during the time period, in concordance with the NABBS estimate of a 2.5% population decrease annually during the 7-year time period (Sauer et al. 2008). Inability to model trend in counts with decision trees may have contributed to the poor performance. The random effect associated with route indicated a positive effect in the core of the breeding range and a negative effect in peripheral regions of the BCR (Figure 11). The relatively high and low values for routes in West Virginia and the northern and southern ends of the BCR suggest that the effect is not well-modeled.

Bart et al. (2004) noted several potential biases with the NABBS survey design. While the hierarchical modeling employed here accounts for biases associated with observer, year, and spatial structure, bias in habitat sampling, survey duration and survey coverage still may prove problematic. The modeling methods used here should be applied to other datasets, specifically off-road surveys, to test if bias in survey location may affect model accuracy and results.

It should also be noted that the results from these models represent patterns in the data and do not infer processes. Many of the variables identified as important to cerulean warbler

presence and abundance may be surrogates for other variables. While inclusion of variables such as latitude and longitude may improve overall model accuracy, the variables can produce artifacts that may not necessarily be ideal for management practices. Additionally, results should be interpreted with the caveat that abundance does not imply habitat quality. Reproductive success is a better measure of quality, and future research should be directed to gaining demographic information for incorporation into spatial models.

## **MANAGEMENT IMPLICATIONS**

The North American Landbird Conservation Plan established a target cerulean warbler population of twice the estimated population of birds present in 1995 (Rich et al. 2004). Because percent forest in the landscape, as well as interior forest, was identified from these models as important to cerulean warblers, increasing the amount of forest throughout the BCR may work toward achieving this population goal. Additionally, the Cerulean Warbler Technical Group has been working to develop a set of guidelines specific to silviculturists and other land managers to address habitat management objectives (Hamel 2006). While variables outlined in large-scale modeling efforts such as mine may not be of a fine enough resolution to completely develop specific silvicultural prescriptions, they may be important in a landscape context. Forest retention and expansion may benefit the species. Additionally, upon field validation, the maps of abundance I have created could be used to develop management objectives to increase populations in areas of low abundance, and to secure populations in locations with higher abundance.

## **LITERATURE CITED**

- ACJV. 2005. Appalachian BCR Concept Plan Version 1.0 July 2005. Atlantic Coast Joint Venture. [Online] [http://www.acjv.org/documents/bcr28\\_concept\\_plan.pdf](http://www.acjv.org/documents/bcr28_concept_plan.pdf).
- Bart, J., K.P. Burnham, E.H. Dunn, C.M. Francis, and C.J. Ralph. 2004. Goals and strategies for estimating trends in landbird abundance. *Journal of Wildlife Management*. 68:611-626.
- Beard, K.H., N. Hengartner, and D.K. Skelly. 1999. Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology* 113:1108-1116.
- Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. *American Scientist* 54:691-692.

- Besag, J., J. York, and A. Mollie. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 54:1-59.
- Best, N. G., R. A. Arnold, A. Thomas, L. A. Waller, and E.M. Conlon. 1999. Bayesian models for spatially correlated disease and exposure data. *Bayesian Statistics* 6:131–156.
- Bibby, C.J., N.D. Burgess, and D.A. Hill. 1992. *Bird Census Techniques*. Academic Press Inc., San Diego, CA.
- Brooks, S.P., and A. Gelman. 1998. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7:434-455.
- Bosworth, S.B. 2003. Cerulean Warbler relative abundance and frequency of occurrence relative to large-scale edge. MS Thesis, West Virginia University, Morgantown.  
[https://etd.wvu.edu/etd/etdDocumentData.jsp?jsp\\_etdId=3116](https://etd.wvu.edu/etd/etdDocumentData.jsp?jsp_etdId=3116).
- Bourg, N.A., W.J. McShea, and D.E. Gill. 2005. Putting a CART before the search: successful habitat prediction for a rare forest herb. *Ecology* 86:2793-2804.
- Breiman, L., J.H. Friedman, R.A. Olshen, and C.G. Stone. 1984. *Classification and regression trees*. The Wadsworth statistics/probability series. Chapman and Hall, New York, NY.
- Buehler, D.A., M.J. Welton, and T.A. Beachy. 2006. Predicting cerulean warbler habitat use in the Cumberland Mountains of Tennessee. *Journal of Wildlife Management* 70:1763-1769.
- Burnham, K.P., and D.R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. 2<sup>nd</sup> edition. Springer-Verlag, New York, NY.
- Choi, Y., H. Ahn, and J.J. Chen. 2005. Regression trees for analysis of count data with extra Poisson variation. *Computational Statistics and Data Analysis* 49:893-915.
- Congalton, R.G. 1991. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of the Environment* 37:35-56.
- Cunningham, R.B., and D.B. Lindenmayer. 2005. Modeling count data for rare species: some statistical issues. *Ecology* 86:1135-1142.
- De'ath, G., K.A. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.
- Dettmers, R., and J. Bart. 1999. A GIS modeling method applied to predicting forest songbird habitat. *Ecological Applications* 9:152-163.
- Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G.



- Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods for predicting species distribution from occurrence data. *Ecography* 29:129-151.
- ESRI. 2006. ArcGIS Version 9.1. 380 New York St, Redlands, CA 92373-8100.
- Federal Register. 2006. Endangered and threatened wildlife and plants; 12-moth finding on a petition to list the cerulean warbler (*Dendroica cerulea*) as threatened with critical habitat. 71:70717-70733.
- Fitzgerald, J. A., W. E. Thogmartin, R. Dettmers, T. Jones, C. Rustay, J. M. Ruth, T. C. Will, and F. R. Thompson III. 2008. Models for conservation planning for terrestrial birds in North America. Pages 593–624 in J. J. Millspaugh and F. R. Thompson III, editors. *Models for planning wildlife conservation in large landscapes*. Academic/Elsevier, Boston, Massachusetts, USA/Amersterdam, Netherlands.
- Franklin, J. 1998. Predicting the distribution of shrub species in Southern California from climate and terrain-derived variables. *Journal of Vegetation Science* 9:733-748.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2004. *Bayesian Data Analysis*. Second Edition. Chapman and Hall, New York, New York, USA.
- Gilks, W.R., and G.O. Roberts. 1996. Strategies for improving MCMC. Pages 89-114 in W.R. Gilks, S. Rochardson, and D.J. Spiegelhalter, editors. *Markov chain Monte Carlo methods in practice*. Chapman and Hall, New York, New York, USA.
- Gilks, W.R, R.S. Rochardson, and D.. Spiegelhalter. 1996. Introducing Markov chain Monte Carlo. Pages 1-20 in W.R. Gilks, S. Rochardson, and D.J. Spiegelhalter, editors. *Markov chain Monte Carlo methods in practice*. Chapman and Hall, New York, New York, USA.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993-1109.
- Gutzwiller, K.J., and W.C. Barrow Jr. 2001. Bird-landscape relations in the Chihuahuan Desert: coping with uncertainties about predictive models. *Ecological Applications* 11:1517-1532.
- Hahn, D.C., and R.J. O'Connor. 2002. Contrasting determinants of abundance in ancestral and colonized ranges of an invasive brood parasite. Pages 219-228 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Hamel, P., D.K. Dawson, and P.D. Keyser. 2004. How we can learn more about the cerulean warbler (*Dendroica cerulea*) *Auk* 121:7-14.

- Hamel, P. 2006. Adaptive forest management to improve habitats for cerulean warbler. Proceedings of Society of American Foresters National Convention 2006.
- Homer C, C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Land-cover Database for the United States. Photogrammetric Engineering and Remote Sensing 70:829-840.
- Keller, C.M.E., J.T. Scallan. 1999. Potential roadside biases due to habitat changes along breeding bird survey routes. Condor 101:50-57.
- Landis, J.R., and G.G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 33:613-619.
- Levin, S.A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943-1967.
- Li, H., and J. Wu. 2004. Use and misuse of landscape indices. Landscape Ecology 19:389-399.
- Link, W.A., E. Cam, J.D. Nichols, and E.G. Cooch. 2002. Of BUGS and birds: a Markov chain Monte Carlo for hierarchical modeling in wildlife research. Journal of Wildlife Management 66:277-291.
- Link, W.A., and J.R. Sauer. 2002. A hierarchical analysis of population change with application to cerulean warblers. Ecology 83:2832-2840.
- McPherson, J.M., W. Jetz, and D.J. Rogers. 2006. Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions – possibilities and limitations. Ecological Modelling 192:499-522.
- O'Brien, C.S., S.S. Rosenstock, J.J. Herver, J.L. Bright, and S.R. Boe. 2005. Landscape-level models of potential habitat for Sonoran pronghorn. Wildlife Society Bulletin 33:24-34.
- O'Conner, R.J., M.T. Jones, D. White, C. Hunsaker, T. Loveland, B. Jones, and E. Preston. 1996. Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States. Biodiversity Letters 3:97-110.
- Ontario Ministry of Natural Resources. 2008. Patch Analyst v4.0. <http://flash.lakeheadu.ca/~rrempe/patch/index.html>.
- Peterjohn, B.G. 2001. Some considerations on the use of ecological models to predict species' distributions. Condor 103:661-663.
- Phillips, S.J., R.P. Anderson, R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.
- Rich, T., C. Beardmore, H. Berlanga, P. Blancher, M. Bradstreet, G. Butcher, D. Demarest, E. Dunn, W. Hunter, E. Iñigo-Elias, J. Kennedy, A. Martell, A. Panjabi, D. Pashley, K.

- Rosenberg, C. Rustay, J. Wendt, T. Will. 2004. Partners in Flight North American Landbird Conservation Plan. Ithaca, NY, Cornell University, Laboratory of Ornithology. 84p.
- R. 2008. The R Foundation for Statistical Computing. Version 2.6.2.
- Riitters, K.H., R.V. O'Neill, and K.B. Jones. 1997. Assessing habitat suitability at multiple scales: a landscape-level approach. *Biological Conservation* 81:191-202.
- Robbins, C.S., J.W. Fitzpatrick, and P.B. Hamel. 1992. A warbler in trouble: *Dendroica cerulea*. Pages 549-562 in J.M Hagan III, and D.W. Johnston, editors. *Ecology and Conservation of Neotropical Migrant Landbirds*. Smithsonian Institution Press, Washington D.C.
- Rosenberg, K.V., S.E. Barker, and R.W. Rohrbach. 2000. An atlas of cerulean warbler populations. Final Report to the U.S. Fish and Wildlife Service, December 2000.
- Royle, J.A. W.A. Link, and J.R. Sauer. 2002. Statistical mapping of count survey data. Pages 625-638 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Sauer, J.R., B.G. Peterjohn, and W.A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* 111:50-62.
- Sauer, J. R., J. E. Hines, and J. Fallon. 2008. The North American Breeding Bird Survey, Results and Analysis 1966 - 2007. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, MD <http://www.mbr-pwrc.usgs.gov/bbs/bbs.html>.
- Scott, J.M., J.H. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson, editors. 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Speigelhalter, D.J., A. Thomas, and N.G. Best. 2000. WinBUGS Version 1.3. User manual. Medical Research Council Biostatistics Unit, Cambridge, UK.
- Speigelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:583-639.
- Thogmartin, W.E., J.R. Sauer, and M.G. Knutson. 2004. A hierarchical spatial model of avian abundance with application to cerulean warblers. *Ecological Applications*. 14:1766-1779.
- Thogmartin, W. E. 2007. Effects at the landscape scale may constrain habitat relations at finer scales. *Avian Conservation and Ecology - Écologie et conservation des oiseaux* 2 (2): 6. [online] URL: <http://www.ace-eco.org/vol2/iss2/art6/>

- Thogmartin, W.E., J.R. Sauer, and M.G. Knutson. 2007. Modeling and mapping abundance of American woodcock across the Midwestern and Northeastern United States. *The Journal of Wildlife Management* 71:376-382.
- Thogmartin, W.E., J.A. Fitzgerald, C.A. Drew, and M.T. Jones. 2008. Conservation design: Where do we go from here? Proceedings of the 4th International Partners in Flight Conference, McAllen, Texas, USA.
- Vayssières, M.P., R.E. Plant, and B.H. Allen-Diaz. 2000. Classification trees: An alternative approach for predicting species distributions. *Journal of Vegetation Science*. 11:679-694.
- Verner, J., M.L. Morrison, and C.J. Ralph, editors. 1986. *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates*. University of Wisconsin Press, Madison, WI.
- Weakland, C.A., and P.B. Wood. 2005. Cerulean Warbler (*Dendroica cerulea*) microhabitat and landscape-level habitat characteristics in southern West Virginia. *Auk* 122:497-508.
- Wiens, J.A. 1989. *The ecology of bird communities*. Vol. 1, Processes and variations. Cambridge University Press, Cambridge.
- Wilds, S.P.V. 1996. Gradient analysis of the distribution of flowering dogwood (*Cornus florida*) and dogwood anthracnose (*Discula destructiva*) in western Great Smoky Mountains National Park. MS Thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC.
- Will, T.C., J.M. Ruth, K.V. Rosenberg, D. Krueper, D. Hahn, J. Fitzgerald, R. Dettmers, and C.J. Beardmore. 2005. The five elements process: Designing optimal landscapes to meet bird conservation objectives. Partners in Flight Technical Series No. 1. <http://www.partnersinflight.org/pubs/ts/01-FiveElements.pdf>
- Wood, P.B., M.P. Strager, J.M. Strager. 2006. Fine-scale forest bird habitat modeling for the mountaintop mining region within the Appalachian Bird Conservation Region. Final Report to the USGS, August 2006.

**Table 1.** Variables used to predict presence/absence and abundance of cerulean warblers in the Appalachian Bird Conservation Region (BCR28). Variables were only used at scales that would be biologically relevant to the focal species.

Variable Description	Scale (m) <sup>a</sup>			Source
	100	300	1000	
Elevation	x	x	x	USGS 30 m National Elevation Dataset
Percent forest in buffer	x	x	x	2001 National Land Cover Dataset
Max size forest patch intersecting buffer	x	x	x	2001 National Land Cover Dataset
PFF score (forest connectivity)		x	x	2001 National Land Cover Dataset
Interior forest		x	x	2001 National Land Cover Dataset
Percent developed in buffer	x	x	x	2001 National Land Cover Dataset
Distance to forested streams			x	NHDplus; National Hydrography Dataset
Latitude	s c a l e l e s s			Breeding Bird Survey
Longitude	s c a l e l e s s			Breeding Bird Survey

<sup>a</sup> Refers to the size of the buffer around each North American Breeding Bird Survey (NABBS) route.

**Table 2.** Confusion matrix from the classification tree for cerulean warbler presence/absence. 104 (randomly chosen) NABBS routes were withheld from model creation for evaluation.

Modeled Group	Actual Count Data		
	Absent	Present	Total
Absent	54	10	64
Present	13	27	40
Total	67	37	104
Accuracy	0.81	0.73	0.78
Kappa	0.53		

**Table 3.** Model averaged posterior distributions of standardized model parameters with 95% credibility intervals. All estimates were included in the creation of a final distribution map for cerulean warbler abundance.

	avg(theta)	2.50%	97.50%
1000 m			
Elevation*	-0.303	-0.556	-0.051
Percent forest in buffer*	0.864	0.455	1.273
Max size forest patch in buffer	0.150	-0.622	0.922
Percent developed in buffer	0.135	-0.070	0.340
Distance to forested stream	0.060	-0.157	0.278
PFF score (forest connectivity)*	-0.193	-0.370	-0.015
Interior forest*	0.225	0.073	0.376
300 m			
Elevation*	-0.171	-0.319	-0.023
Percent forest in buffer*	0.545	0.340	0.750
Max size forest patch in buffer	0.058	-0.458	0.574
Percent developed in buffer*	0.118	0.001	0.235
PFF score (forest connectivity)	-0.126	-0.304	0.052
Interior forest	0.068	-0.253	0.389
100 m			
Elevation*	-0.060	-0.119	-0.002
Percent forest in buffer*	0.196	0.141	0.251
Max size forest patch in buffer	-0.002	-0.210	0.206
Percent developed in buffer*	0.080	0.030	0.130

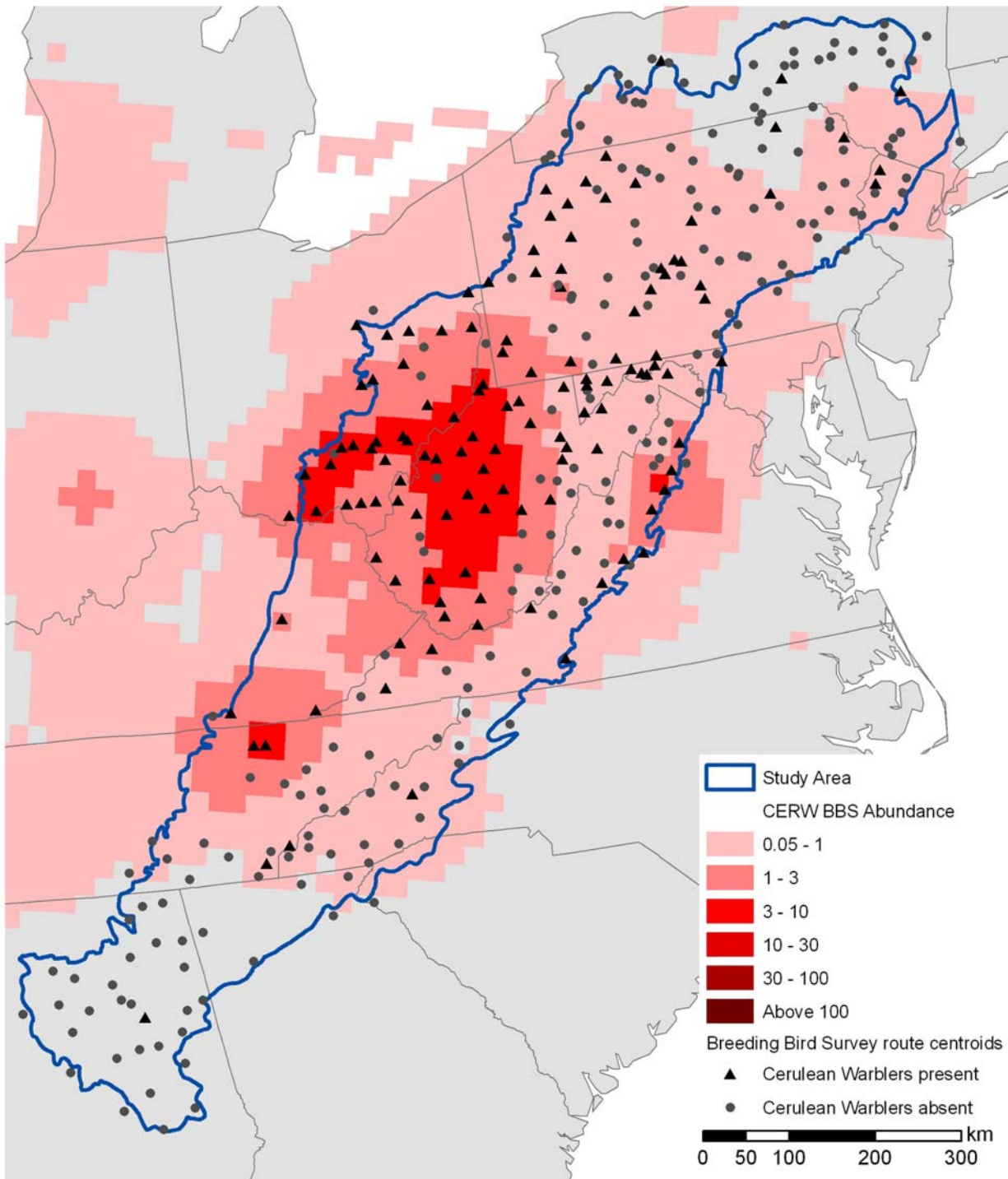
\* indicates a credible covariate

**Table 4.** DIC values and model weights for final models at each scale for hierarchical spatial analyses.

model	DIC	Deviance	GOF	$\Delta$ DIC	$w_i$	evidence ratios
1000 m model	2277.12	2020.27	0.5232	0.00	0.5393	1.0000
300 m model	2278.10	2031.38	0.4978	0.98	0.3304	1.6323
100 m model	2279.96	2031.44	0.5173	2.84	0.1304	4.1371

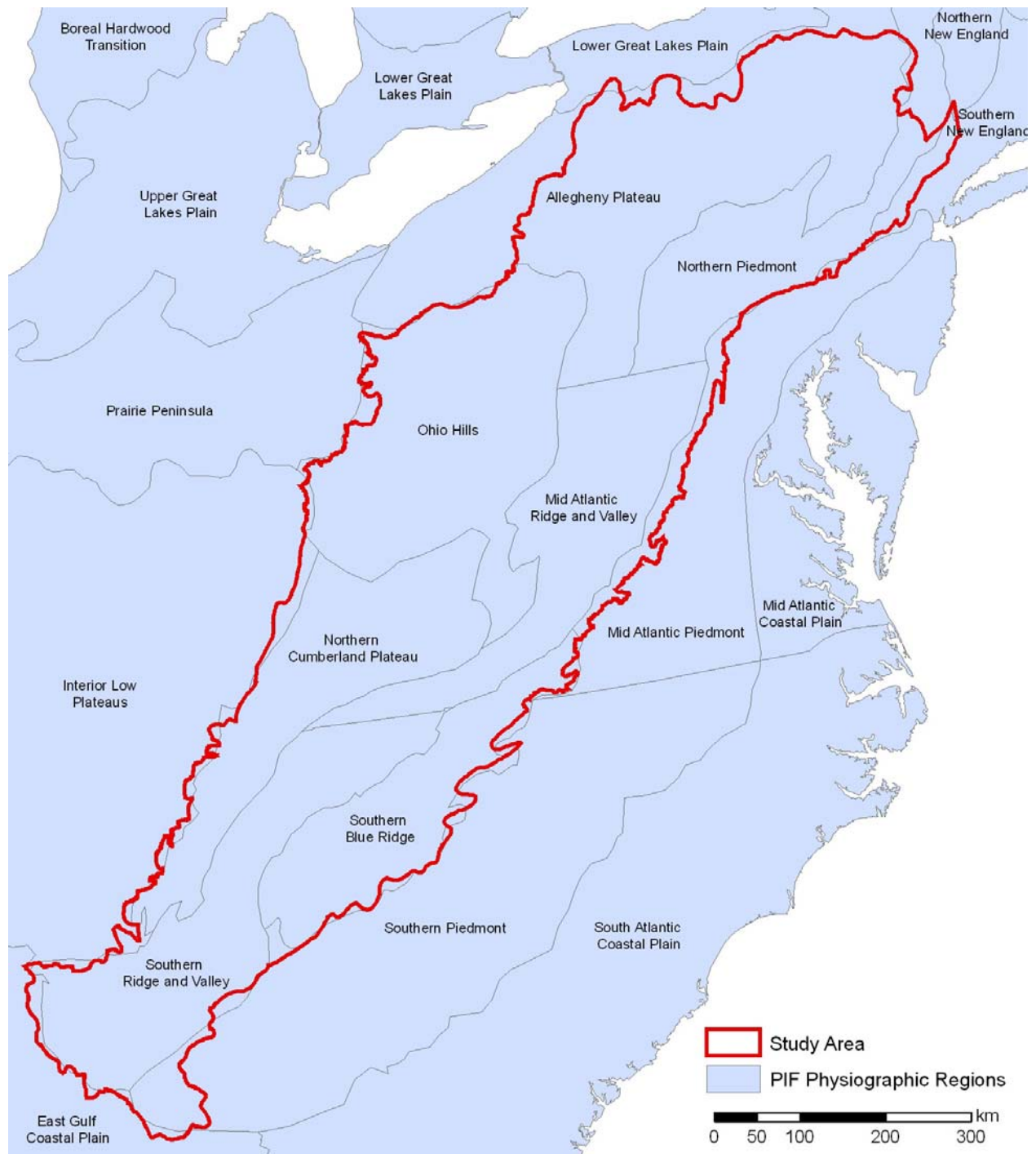
**Table 5.** Standard deviation associated with random effects. Larger standard deviation values represent larger variation explained by the random effect.

node	mean	2.50%	97.50%
sdnoise	0.396	0.274	0.519
sdobs	0.607	0.278	0.979
sdrte	3.110	2.447	3.842
sdyears	0.153	0.031	0.472
yeareffect[1]	0.120	-0.163	0.521
yeareffect[2]	-0.047	-0.359	0.236
yeareffect[3]	-0.055	-0.355	0.191
yeareffect[4]	-0.073	-0.382	0.157
yeareffect[5]	-0.040	-0.356	0.227
yeareffect[6]	0.054	-0.243	0.397
yeareffect[7]	0.037	-0.325	0.418
yearplustrend[1]	0.281	-0.137	0.755
yearplustrend[2]	0.087	-0.374	0.492
yearplustrend[3]	0.053	-0.404	0.437
yearplustrend[4]	0.007	-0.440	0.360
yearplustrend[5]	0.014	-0.407	0.376
yearplustrend[6]	0.081	-0.274	0.477
yearplustrend[7]	0.037	-0.325	0.418

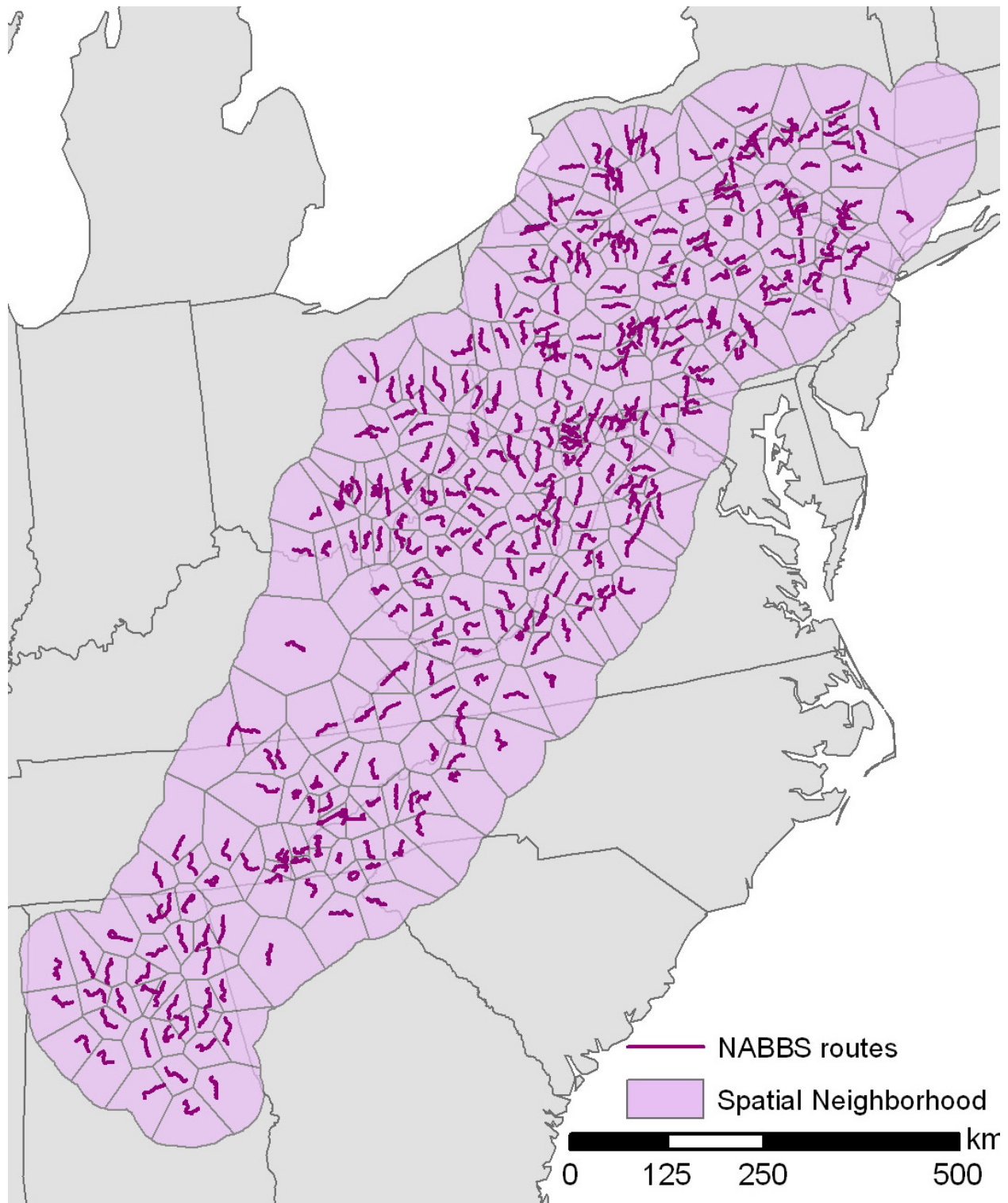


**Figure 1.** Breeding range of the cerulean warbler as derived from an interpolation of Breeding Bird Survey (NABBS) counts from 1961 – 2006 (Sauer et al. 2008). Black triangles represent centroids for NABBS routes where at least one Cerulean Warbler was detected between 1998 and 2004. Grey dots represent routes where no Cerulean Warblers were detected during the time period.

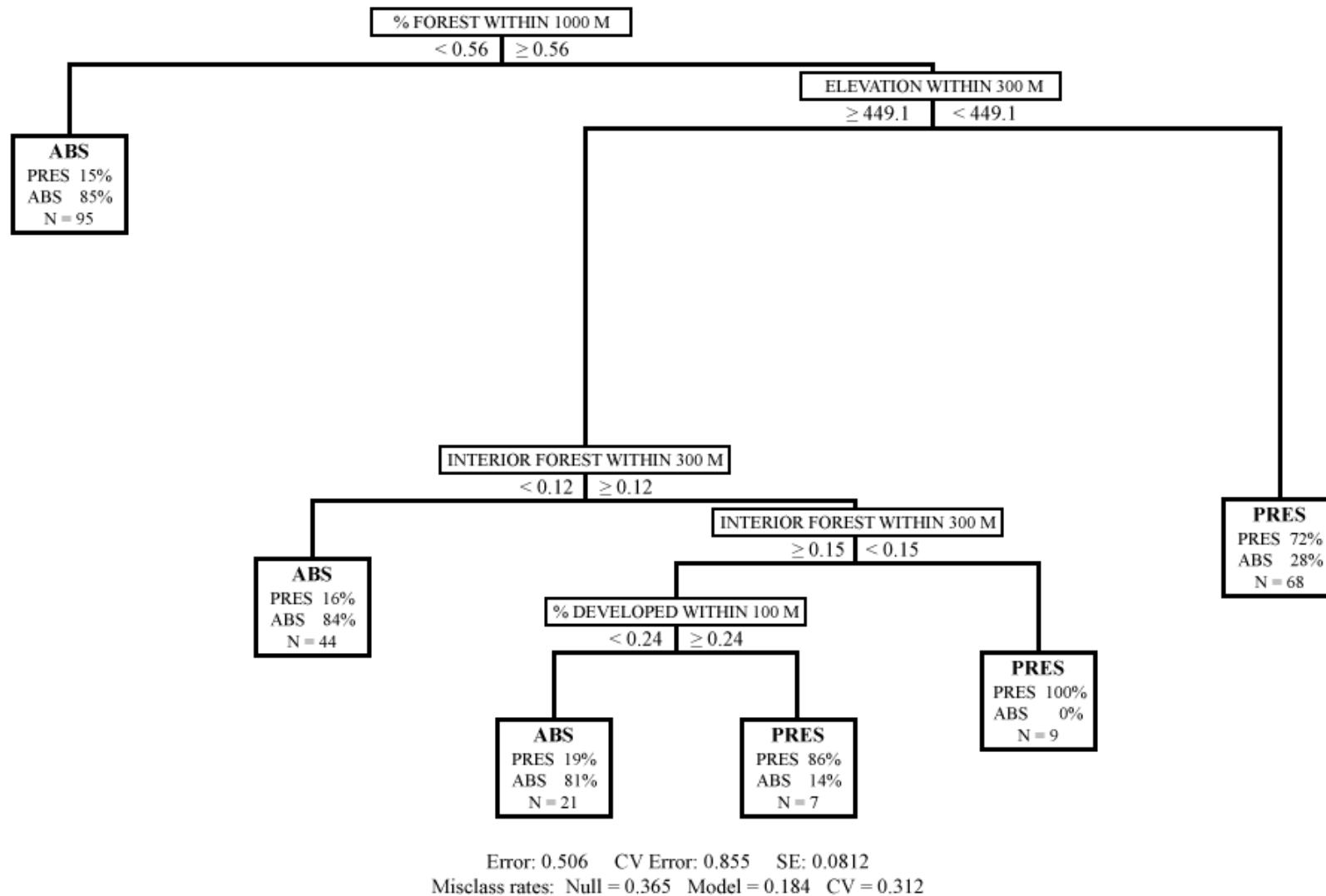




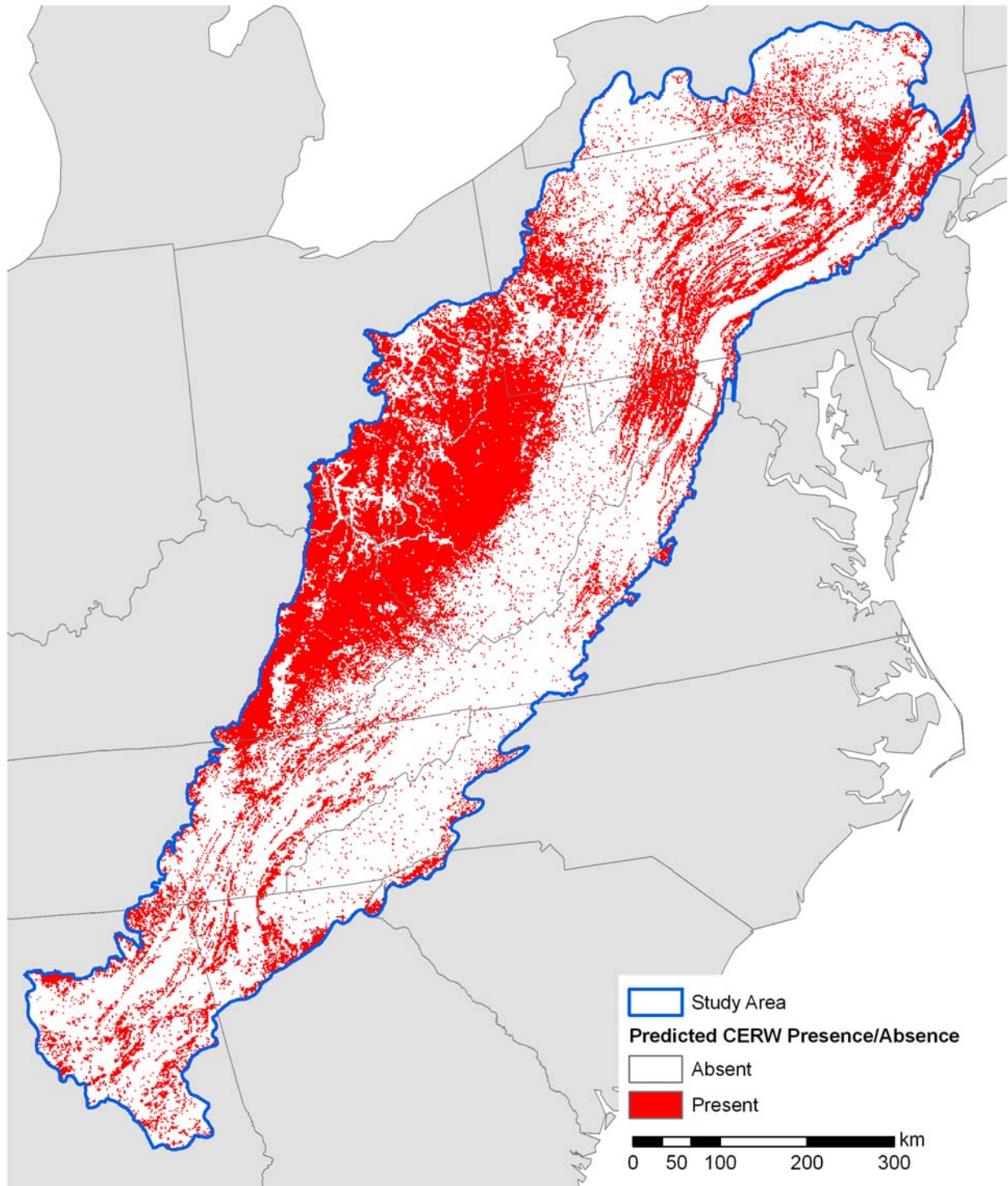
**Figure 2.** Partners in Flight physiographic regions and study area boundary.



**Figure 3.** Neighborhood structure for NABBS routes within the Appalachian Mountains Bird Conservation Region (BCR28). Only first order neighbors were used in the adjacency matrix to model the spatial structure of counts. Neighborhood structure was created by tessellating North American Breeding Bird survey routes (NABBS;  $n=348$ ) within BCR28.



**Figure 4.** Classification tree model for cerulean warbler presence (PRES)/absence (ABS). Values below each node represent the number of routes assigned to each associated category.

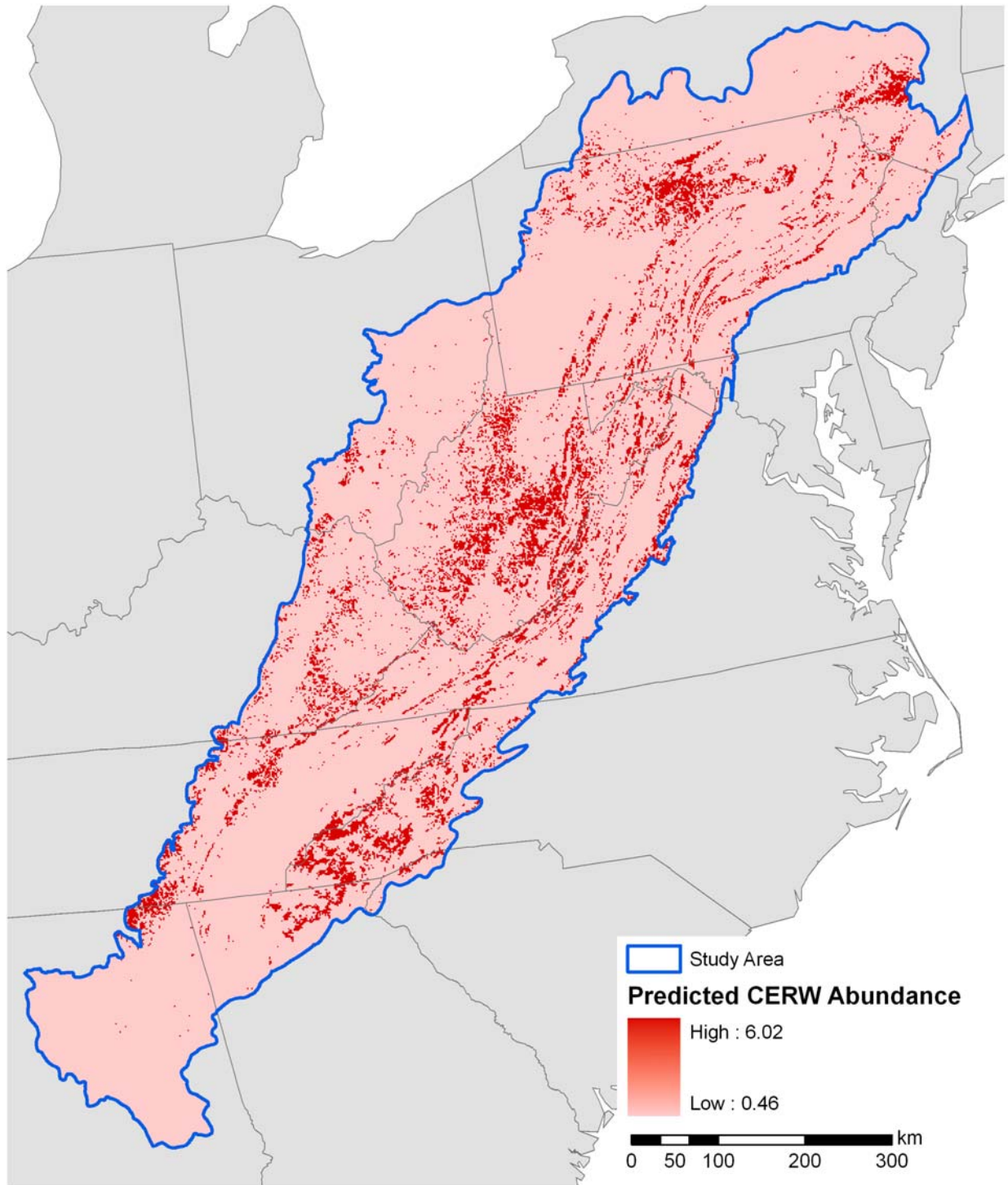


**Figure 5.** Predicted presence/absence of cerulean warblers within the Appalachian Bird Conservation region. Results are based on classification tree analysis and represent sum of cerulean warblers detected on a NABBS survey route (n=348 routes).



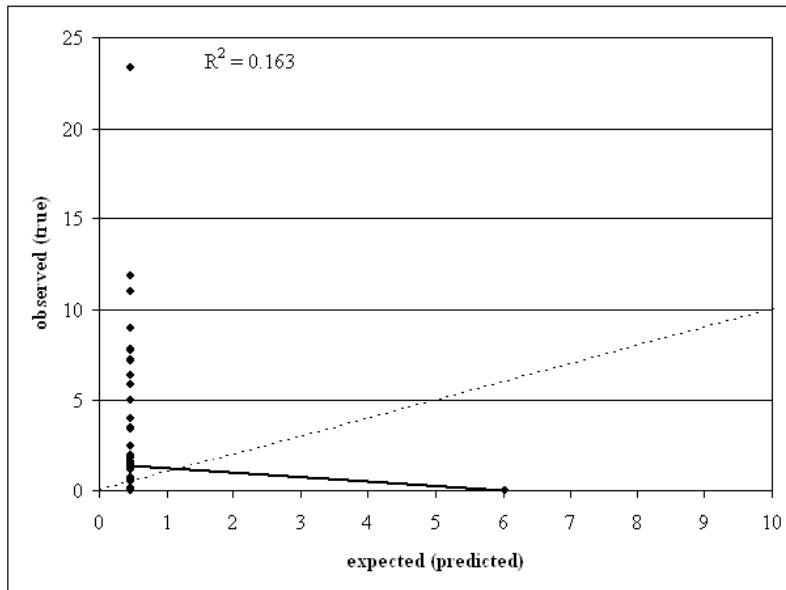


**Figure 6.** Regression tree model for cerulean warbler abundance. Each node ends with the estimated abundance of cerulean warblers and the number of routes classified by the node.

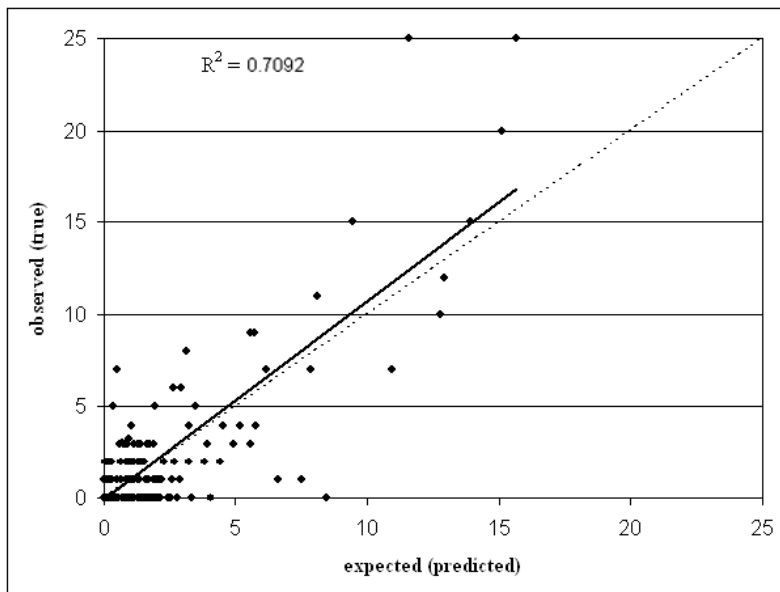


**Figure 7.** Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on regression tree analysis and represent sum of cerulean warblers detected on a NABBS survey route (n=348 routes).

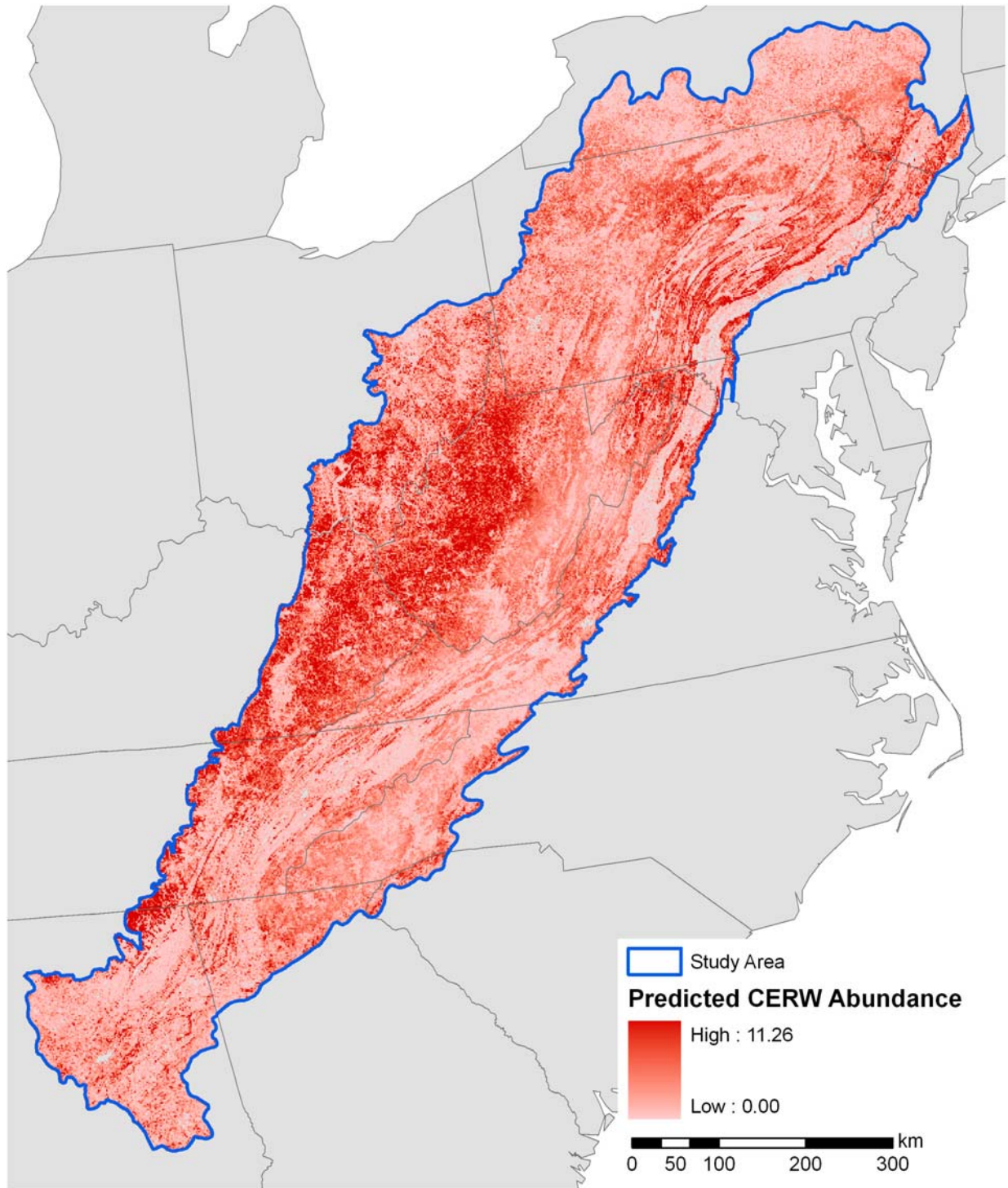
### a. Regression tree model



### b. Hierarchical spatial model

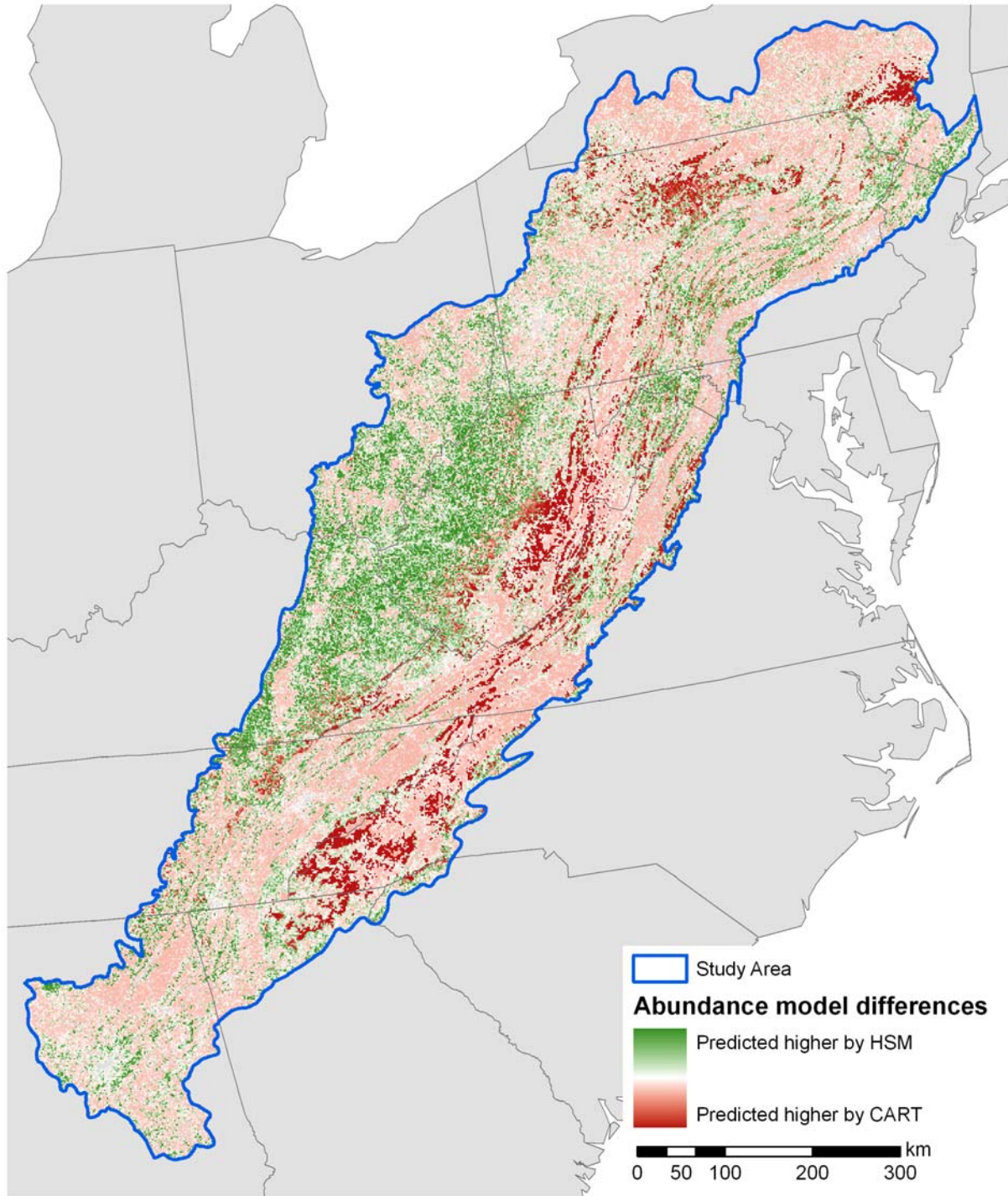


**Figure 8.** Observed maximum North American Breeding Bird Survey (NABBS) counts (a:  $n = 104$  validation routes) and observed yearly NABBS counts (b:  $n=594$ ) during 1998 – 2004 for cerulean warblers in the Appalachian Bird Conservation Region (BCR28) compared to expected (predicted) counts from the best regression tree and hierarchical spatial model of cerulean warbler abundance. Observed counts were withheld from model construction. Dashed lines represent a one-to-one correspondence between observed and predicted counts. We fitted observed versus expected counts with simple linear regression (solid line). Each point may represent more than one route or stop.

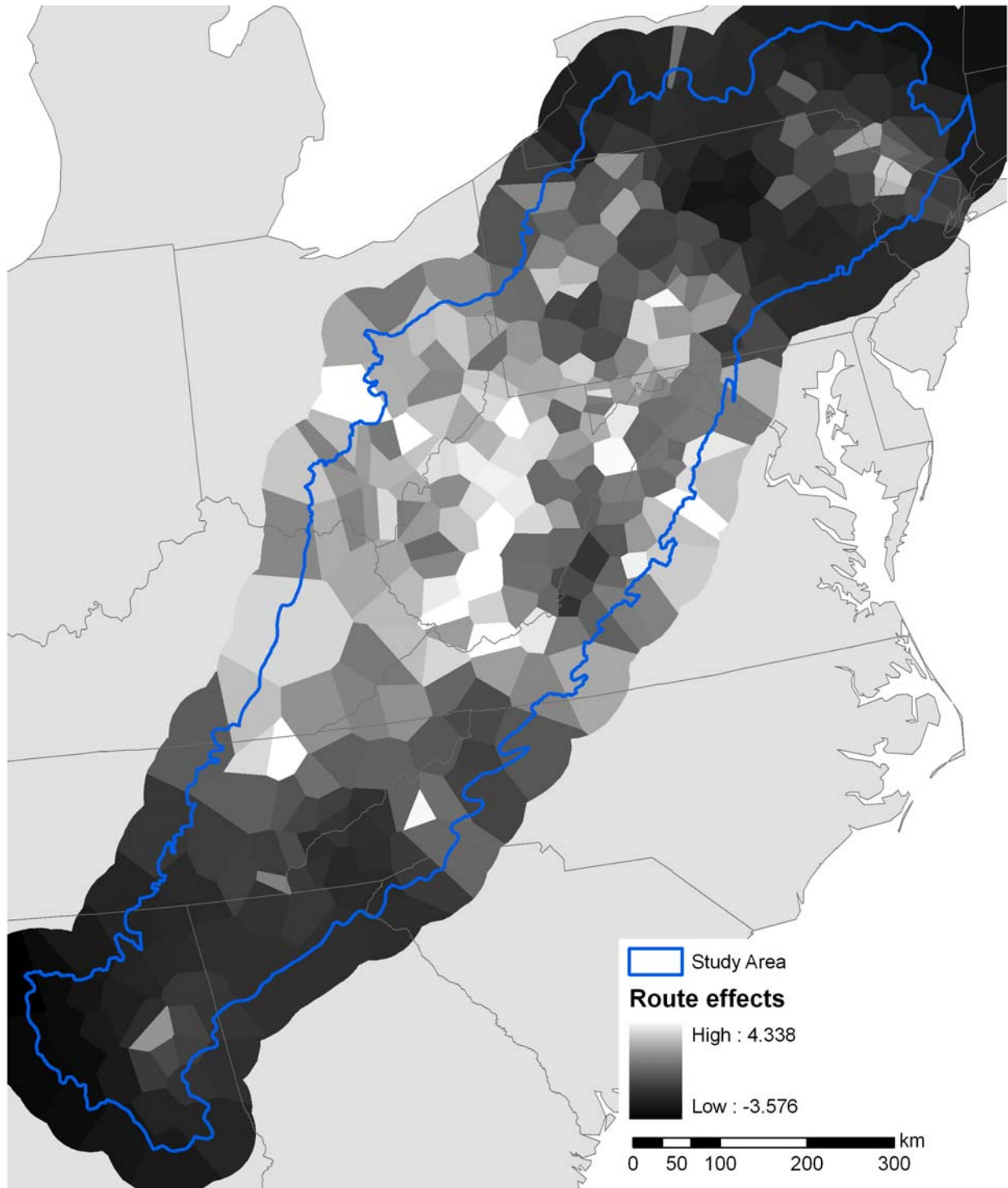


**Figure 9.** Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on a hierarchical spatial analysis and represent sum of cerulean warblers detected on a NABBS survey route (n=348 routes).





**Figure 10.** Variation between predicted abundance maps created by decision trees and hierarchical spatial models. Green cells indicate areas where cerulean warblers are predicted in higher abundances by hierarchical spatial models. Red cells indicate areas where cerulean warblers are predicted in higher abundances by decision tree models. Maps were created individually using results from statistical models and compared by assessing the mathematical difference between raster cells in ArcGIS 9.2.



**Figure 11.** Mapped route effects (n=348 NABBS routes) for the 300 m hierarchical spatial model. Positive numbers indicate an increasing effect of route on abundance, while negative numbers indicate a decreasing effect of route on abundance.

## **Chapter 3**

# **COMPARISON OF BREEDING BIRD SURVEY ROUTE- AND STOP-LEVEL SPATIAL MODELS WITH APPLICATION TO CERULEAN WARBLERS**

**Formatted in the style of *Ecological Applications***

## ABSTRACT

One of the greatest uncertainties facing wildlife-habitat relationship models is scale. The definition of scale has been used differently among ecologists, however three distinct components should be identified: 1) the resolution of species distribution data; 2) the resolution of habitat variables; and 3) the extent of the study area. Resolution of habitat data, and study extent have recently received much attention, but the resolution of species distribution data and how it relates to the other components of scale has received little attention. Microhabitat data may only be appropriate for site-specific analyses. Conversely, patch fragmentation metrics may only be applicable to regional analyses. Here I present a comparison of two model types, differing in the resolution of the species distribution data. I used North American Breeding Bird Survey (NABBS) counts with a suite of explanatory variables to predict presence and abundance of cerulean warblers (*Dendroica cerulea*) in the Appalachian Mountains Bird Conservation Region (BCR28). Decision trees were created for route-level and stop-level analyses of presence and abundance. Additionally, output maps have typically been resolved to the resolution of the environmental spatial datasets with little attention given to the scale at which the predictions represent. Using the modeling results, predictive distribution maps were created for cerulean warblers with appropriate resolutions for each model group. Route-level decision trees performed better than stop-level models for predicting both presence and abundance of cerulean warblers. Similar to raw NABBS distribution data, cerulean warblers were predicted to occur in highest concentrations in the central portions of the BCR. Poor performance of stop-level models may result from a mismatch of resolution of environmental data to species survey data, or lack of important environmental covariates at the stop-level scale. The results of this study highlight the importance of correctly matching the resolution of the species distribution data to the resolution of environmental covariates and the extent of analysis. Additionally, the results and relationships highlighted here may serve to direct management and monitoring for the cerulean warbler.

**KEY WORDS** Classification and regression trees, cerulean warbler, count data, *Dendroica cerulea*, North American Breeding Bird Survey, resolution, scale, spatial modeling.

## INTRODUCTION

The need for accurate models that identify habitats for wildlife conservation and restoration is becoming increasingly common. However, uncertainty in model design is still an important factor that should be taken into consideration *a priori*. A fundamental uncertainty concerning wildlife habitat relationship models (WHR) is scale. Tobalske (2002) identified three important components of scale that can characterize WHR models: 1) the resolution of the species distribution data; 2) the resolution of the habitat variables; and 3) the extent of the study area. However, the definition of scale is often confusing and used differently among ecologists.

It is becoming evident that scale (resolution and extent of the data) is an important factor in ecological modeling. But inappropriate use of these components of scale may affect an incomplete or misleading interpretation (Bowyer and Kie 2006). Particular emphasis has been given to examining analysis extent for ecological studies to identify scales at which ecological processes may occur (Scott et al. 2002). Additionally, resolution of environmental data has received attention in recent years (Divine et al. 2000, Scott et al. 2002). However, resolution of species occurrence data and how it relates to the resolution of environmental variables has not been thoroughly addressed.

Real-world applicability of variables and data used for specific levels of analysis must be taken into consideration. Microhabitat data may only be appropriate for site-specific or local-level analyses. Conversely, patch fragmentation metrics may only be applicable to wider-extent or landscape-level analyses (Zabel et al. 2002). Matching the right suite of variables with the appropriate scale is an important step for elucidating patterns in ecological data. Land managers and biologists are constantly faced with prioritization questions. Methods and tools to efficiently speed up the analyses and prioritization, while properly considering scale, will be a boon to wildlife management and conservation.

Classification and regression trees (CART) have successfully been used to predict habitat use and distributions for a variety of taxa including songbirds (O'Conner et al. 1996, Wood et al. 2006b), mammals (O'Brien et al. 2005), and plants (Bourg et al. 2005). CART, a data-mining technique, operates by recursive partitioning thereby creating decision trees that are similar in concept to dichotomous trees (Vayssières et al. 2000). Terminal nodes of the tree result in a classification of the response variable. Decision trees have performed as well as other techniques such as logistic regression, ANOVA, and linear regression, yet may be better at

revealing patterns in datasets (Vayssières et al. 2000, De'ath and Fabricius 2000, Bourg et al. 2005). Results of CART models must be interpreted with the caveat that they are largely exploratory in nature and only return estimates of correlation, thus, results are subject to the normal limitations of correlation analysis, specifically that correlation does not ensure causation (Hahn and O'Connor 2002). Yet trees have many advantages including the flexibility to handle a broad range of response types, ease and robustness of construction, ease of interpretation, and the ability to handle missing values in both response and explanatory variables (De'ath and Fabricius 2000). While other techniques may better handle nuisance effects such as observer bias and seasonal and temporal differences, they are often much more difficult to implement than CART and can require enormous computing power.

Typically, statistical models are not the final product used in management applications. Modeling results are often combined with spatial layers to create predicted species distribution maps as well as maps of potential habitat suitability. Scale (specifically, resolution) has been neglected at this stage of the process. It is common for output maps to match the resolution of the input variables, but data collection and modeling methods should be taken into account for final output resolution. Ultimately, map resolution should be a function of the data extent. For example, species' count data averaged over a survey route or transect might have an output cell size for predictions comparable to the extent of the route or even the scale at which environmental covariates are measured (e.g., percent forest within 1,000 m of survey route).

Here, I compare two model types, differing in the resolution of the species distribution data. To apply the models, I predicted presence/absence and relative abundance for the cerulean warbler (*Dendroica cerulea*) from North American Breeding Bird Survey (NABBS) data for the Appalachian Mountains Bird Conservation Region (BCR28) in the eastern United States. The cerulean warbler, a small (11.5 cm, 8-10 g) Neotropical migrant that breeds in mature deciduous forest, has its core range in the central Appalachians (Figure 1). A rangewide population decline has made it a focal species of research and management priority (Robbins et al. 1992, Hamel et al. 2004). The NABBS is one of the longest running and most important monitoring projects for avian conservation in North America (Sauer et al. 2008). The 43-year program has been used to determine trends in species' populations over time (Sauer et al. 2003, Sauer et al. 2008), and more recently, it has been used to determine spatial patterns in counts (Thogmartin et al. 2004). Because of the lack of stop coordinates for the 50 survey locations on each route, spatial analyses

have been limited to a route-level resolution even though cerulean warblers respond to fine-scale habitat attributes (e.g. slope position, aspect) better measured at the stop-level (Weakland and Wood 2005, Wood et al 2006a). However, recent efforts to georeference all stop locations will allow finer-resolution analyses to be conducted. For this chapter, I 1) determined and compared the predictive accuracy of each set of models, and 2) created predictive distribution maps for cerulean warblers with appropriate resolutions for each model group. I hypothesize that the stop-level models will better predict cerulean warbler abundance than route-level models, because of the added level of detail.

## METHODS

### *Study area*

The Appalachian Bird Conservation Region (BCR28) comprises >426,000 km<sup>2</sup> throughout 15 states. The portion that overlaps Massachusetts and Connecticut was removed prior to analyses (Figure 1); because of its thin shape, it does not completely contain any North American Breeding Bird Survey (NABBS) routes. The study region intersects or contains 15 Partners in Flight (PIF) physiographic regions (Figure 2).

The study region is dominated by oak-hickory (*Quercus* and *Carya* spp.), mixed-mesophytic, and northern hardwood forest types (ACJV 2005). Topography consists of mostly mountainous regions throughout the Appalachian Mountains with a terrain of highly convoluted ridges and valleys (ACJV 2005). Low-elevation flat portions of the BCR often are in agricultural use (ACJV 2005). Other types of habitat include high elevation red spruce (*Picea rubens*) forests, wetlands, scrublands, and grasslands. The diverse habitat types are partly a result of the large range in latitude, longitude, and elevation (0-1,981 m above sea level).

### *Data sets*

Models were constructed using NABBS data. The NABBS is a long-term roadside survey initiated in 1966 to monitor bird population trends within North America (Sauer et al. 2008). Surveys are completed annually during May and June along randomly assigned roadside routes. Each route contains 50 stops spaced approximately 0.8 km along the route. Observers conduct a 3 minute point count at each stop and record all birds seen or heard within a 402 m radius. Surveys begin 30 minutes before local sunrise and take approximately 5 hours to complete. When possible, individual routes are surveyed by the same observer each year and

under suitable weather conditions (i.e., low wind, low-precipitation, high visibility). The protocol is designed to maintain a level of consistency and minimize variability in sampling methods so variations in bird population trends can be detected over time.

I used count data from NABBS routes within BCR28 that had stop-level coordinates available and were active between 1998 and 2004 ( $n = 115$ ) to develop and validate models. I randomly selected 30% of the data to be withheld for validation. Models were split into two categories representing different resolution of data: route-level and stop-level. For route-level models, each route was assigned the 7-year mean of the sum of counts from the 50 stops. Stop-level models used the actual recorded abundance of cerulean warblers at each stop ( $n=5,355$  on 115 routes) averaged for the 7-year period. I modeled count data collected between 1998 and 2004, a 7-year window that coincides with the time period of the 2001 National Landcover Dataset (NLCD, Homer et al. 2004).

I reviewed previous modeling efforts for Cerulean Warblers (Dettmers et al. 1999, Thogmartin et al. 2004, Buehler et al. 2006, Wood et al. 2006b), and relied on expert opinion to select a parsimonious suite of environmental covariates *a priori* that seemed biologically significant to the focal species (Table 1). These environmental covariates were derived from moderate-resolution remotely-sensed data with a resolution of 30 m x 30 m and primarily represented land cover. Most topographic variables (slope, aspect, and relative slope position) were not included for route-level analyses because they were not appropriate for use over large geographic areas. Topographic variables such as aspect will encompass most values over a large geographic area, thus averaging to an uninformative orientation. Additionally, I did not include climatic variables for either set of models to focus on variables that could more easily be implemented in management activities. Metrics were calculated within each of three buffers (100 m, 300 m, and 1,000 m) around each route and stop using ArcGIS® version 9.1 (ESRI 2006) and the Spatial Analyst (ESRI 2006) and Patch Analyst© version 4 (Ontario Ministry of Natural Resources 2008) extensions. Buffer sizes were chosen to represent 3 unique scales relevant to cerulean warbler biology (specifically, extent or level of analysis); immediate resource area, local, and regional, respectively. The smallest extent is comparable to the nest site and immediate vicinity, the habitat most used by a territorial adult. The local extent would represent the maximum territory size, and regional extent is an area over which the species may select breeding habitat.



The USGS National Elevation Dataset (NED) was used to calculate topographic variables (elevation, aspect, slope and relative slope position). Aspect was transformed according to Beers et al. (1966). I used ArcInfo Workstation (ESRI 2006) to calculate relative slope position, a continuous metric that classifies cells based on the distance of a particular point from the nearest slope bottom to the nearest ridge, as defined by surface flow hydrological analysis using a filled digital elevation model (DEM, Wilds 1996). Land cover metrics were derived from the National Landcover Dataset (NLCD, Homer et al. 2004). Distance to forested streams was calculated from a combination of NLCD and National Hydrography Dataset (NHDPlus; 1:100,000 scale). Additionally, I calculated a moisture index ranging from 0 (dry) to 18 (wet) (Anderson et al. 1998). While moisture may not directly affect cerulean warbler habitat use, it influences forest productivity and local vegetative variables' occurrence and distribution that would directly impact the species (Webster and Jenkins 2005). Because I selected a parsimonious suite of environmental covariates in the models and the data were restricted to a 30 m cell size resolution, some variables important to cerulean warblers may have been omitted. To help accommodate this, I included latitude and longitude in the modeling as a surrogate. This also may help to address some of the geographic patterning in the dataset (Franklin 1998).

#### *Decision tree models*

I used CART to model presence/absence and relative abundance for cerulean warblers. Decision trees were created from the route- and stop-level data sets, each for two separate analyses: (1) classification tree analysis of presence/absence and (2) regression tree analysis of abundance. Analyses were performed in R (R 2008) using the MVPART and CARET packages. I used multiple (50) 10-fold cross-validation to create “pruned” or optimal tree sizes. During each of the 50 runs, the data are divided into 10 subsamples. Each subsample is used once for training and validation purposes (Vayssieres et al. 2000, Bourg et al. 2005). Overgrown trees can reduce the misclassification error of the data set used for construction, but perform poorly on independent validation data. In addition to the multiple cross-validation, I chose final tree sizes that were within 1 standard error of the minimum cross-validation (Breimen et al. 1984). Data withheld from model construction were used to validate model performance.

### *Model evaluation*

I withheld a randomly chosen 30 percent of the data to validate the final models. This amounted to 35 routes and 1607 stops. For classification trees, I calculated the misclassification rate based on the original model development points, confusion matrices, and Cohen's kappa statistic. I also calculated an overall error rate based on the reserved validation data points. The kappa statistic is a measure of the correspondence between model predictions and actual observations and is more stringent than simply using the overall error rate (Congalton 1991). For regression trees, I evaluated model performance using simple linear regression to test the predicted set of values against the actual observations.

### *Mapping*

I used the final models to create distribution maps for cerulean warblers in BCR28. I created spatial maps of cerulean warbler presence/absence and relative abundance for each model group. Results from decision trees were incorporated into Spatial Analyst (ESRI 2006) as a set of nested conditional statements. Data layers were first resampled to represent the scales at which each variable was measured. Grids were resampled using neighborhood statistics with a moving window corresponding to each scale (Wood et al. 2006b).

To properly represent the appropriate scale at which predictions should be interpreted, I resampled output grids. For route-level models, grids were resampled to a cell size of 1,566.5 m x 1,566.5 m to represent the listening radius over 50 stops (area =  $50\pi * 125 \text{ m}^2$ ). For stop-level models, grids were resampled to a cell size of 221.5 m x 221.5 m to directly represent the putative listening radius (125 m; area =  $\pi * 125 \text{ m}^2$ ) for cerulean warblers, as suggested by Partners in Flight (Rosenberg and Blancher 2005). Variation between route- and stop-level maps was compared by assessing the mathematical difference between overlapping raster cells.

## **RESULTS**

### *Route-level models*

Both classification and regression trees for route-level NABBS data performed better than stop-level decision trees. The final classification tree for presence/absence had 3 splits and 4 terminal nodes, with only land cover (percent forest and maximum size forest patch) variables represented (Figure 3). Two of the three scales of analysis were represented (regional and immediate; 1000 m and 100 m, respectively), as well as latitude. Cross-validation with test data

resulted in a model with an overall classification accuracy of 77% (Table 2) and a moderate kappa score (0.54). Classification accuracy of presence points was 85%, while accuracy for absence points was 73%. Cerulean warblers were predicted to occur from northern-most portions of the BCR south through Tennessee (Figure 4). Because latitude was used in the final model, cerulean warblers were predicted to be completely absent from the southern third portion of the BCR.

The final regression tree for abundance had 1 split and 2 terminal nodes (Figure 5). Similar to the classification tree, percent forest was used in discriminating the data. Areas with highest abundance occurred in Pennsylvania, North Carolina, and West Virginia (Figure 6). Data withheld from model construction suggested moderate correspondence between actual and estimated counts ( $R^2 = 0.55$ , Figure 7a).

#### *Stop-level models*

Stop-level classification and regression trees both performed more poorly than their route-level counterparts. The final classification tree for presence/absence had 5 splits and 6 terminal nodes. Stop-level models contained both landcover and topographic variables (maximum size forest patch and percent forest, and moisture and slope, respectively; Figure 8). Two of the three scales of analysis were represented (immediate and local; 100 m and 300 m, respectively) as well as longitude. Cross-validation with test data resulted in a model with a high overall classification accuracy of 95% (Table 3) but a low kappa score (0.22). The high overall accuracy is somewhat misleading, as the model is not much improved from the null model. The kappa score is a better reflection of the model performance. Classification accuracy of presence points was 15%, while accuracy for absence points was 99%. Cerulean warblers were predicted to occur in the central portion of the BCR, primarily in West Virginia and Kentucky (Figure 9).

The final regression tree for abundance had 4 split and 5 terminal nodes (Figure 10). Similar variables as in the route-level regression tree were used to discriminate the data. Though all predicted abundances were low, the distribution of highest predicted abundances (Figure 11) overall agreed with the map for presence/absence (Figure 9). Cerulean warblers were predicted in slightly higher abundances in West Virginia, Kentucky, and North Carolina. Data withheld from model construction suggested poor correspondence between actual and estimated counts ( $R^2 = 0.23$ , Figure 7b). Both stop and route-level models predicted more cerulean warblers than actually occurred on the landscape (Figure 7).

## DISCUSSION

As wildlife management agencies face increasing conservation issues and decreasing budgets, the ability to collect the least amount of data necessary, perform quick analyses, and arrive at proper management decisions becomes more important. Though many programs like the North American Breeding Bird Survey collect species survey data annually at thousands of specific points all across the country, the data have historically been analyzed at a route or regional level. Finer resolution survey data should be paired with fine-resolution land cover, climate, and topography data for best results. How to properly match survey data and environmental data is still a question that remains largely unanswered and probably varies among species, guilds, and even broader groups of organisms.

In this modeling study, route-level models performed better than stop-level models, contrary to the hypothesis I made prior to analysis. For classification trees, overall accuracy was higher for the stop-level model than route-level. However, this accuracy level can be misleading. The kappa score was much lower for the stop-level models, indicating a poor fit. The high accuracy of absence points, and the vast number of absences in the data result in a high overall accuracy. However, it may be more important to have the ability to accurately predict the location of species (presence locations). The large decrease in accuracy of predicted presence locations from 85% to 15% (route- and stop-level models respectively) provides more insight into the differences in kappa.

Performance of regression trees was similar to that of classification trees. Route-level models performed moderately well, suggesting good correspondence between actual and estimated counts ( $R^2 = 0.55$ , Figure 7). However, stop-level models were not able to fit the data as well. The low correspondence between actual and estimated counts ( $R^2 = 0.23$ , Figure 7) may have multiple explanations. One possible explanation is that the variables included in the modeling were not appropriate for the extent and resolution of the survey data. While the topographic and vegetative variables I included may have explained route or regional trends well, they may not be as important at smaller scales. Additionally, the environmental data available may not be collected at a fine enough resolution to explain the patterns in stop-level counts; microhabitat data, which would likely result in more accurate predictions have not been collected at NABBS stop locations. Furthermore, cerulean warblers have been shown to prefer

forested areas with canopy gaps (Perkins 2006). Currently, there is no moderate-resolution dataset at the BCR extent to represent this variable. Lack of this important covariate in the stop-level models may be contributing to the poor predictive performance.

The majority of models included either latitude, longitude, or both in final decision trees. This suggests that the environmental covariates did not sufficiently explain the spatial structuring of counts. There may be variables not included in the modeling that are yet to be identified as important to cerulean warblers, or not yet available as a spatial dataset (e.g., canopy). Moisture and percent forest were identified in several of the models. Like latitude and longitude, moisture may be acting as a surrogate for some microhabitat variable that is important to this bird species. Moisture has been identified in past modeling efforts as important to cerulean warblers (Dettmers et al. 1999, Buehler et al. 2006). More research should be aimed at identifying vegetative variables influenced by moisture that may be important for cerulean warblers and other wildlife species.

Route-level models predicted cerulean warblers to be present in more areas than did stop-level models (Figure 12). This could be an artifact of the way the count data are appended to the route. For instance, cerulean warblers may be absent on 49 of the 50 NABBS stop locations; however, if at least one stop contained cerulean warblers, then the route is labeled as “present.” While the majority of habitat in the area may be unsuitable for the species, it is regarded as being positive in the modeling process. Abundance models were equally skewed, with higher counts predicted by route-level models (Figure 13).

Maps from modeling results have typically been resolved to the resolution of environmental data with little regard for the methods used for the species surveys, or for the resolution (e.g., point, route, or region) of the response data. Ultimately, these prediction maps are used to identify areas where the species may be occurring and to influence management decisions. There may also be legal ramifications to predictive maps; modeling results are often incorporated into management plans and petitions to include wildlife on threatened and endangered lists. Predictions should not be more finely resolved than the data and methods used to create them. Though the patterns are often similar, details are quite variable between the resolutions (Figure 14).

This study was the first to my knowledge that compared predictive accuracy of route- and stop-level NABBS count data. Using current environmental variables, route-level models more

accurately predict cerulean warbler occurrence and abundance throughout the Appalachian Mountains Bird Conservation Region. Further research should be aimed at modeling survey data using different resolutions of environmental data to find appropriate matches for modeling.

## **LITERATURE CITED**

- ACJV. 2005. Appalachian BCR Concept Plan Version 1.0 July 2005. Atlantic Coast Joint Venture. [Online] [http://www.acjv.org/documents/bcr28\\_concept\\_plan.pdf](http://www.acjv.org/documents/bcr28_concept_plan.pdf).
- Anderson, M.G., M.D. Merrill, and F.B. Biasi. 1998. Connecticut River Watershed: Ecological communities and neo-tropical migratory birds. Final Report Summary. The Nature Conservancy, Eastern Conservation Science, Boston Satellite Office, Boston, MA.
- Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. *American Scientist* 54:691-692.
- Bourg, N.A., W.J. McShea, and D.E. Gill. 2005. Putting a CART before the search: successful habitat prediction for a rare forest herb. *Ecology* 86:2793-2804.
- Bowyer, T., and J.G. Kie. 2006 Effects of scale on interpreting life-history characteristics of ungulates and carnivores. *Diversity and Distributions* 12: 244-257.
- Breiman, L., J.H. Friedman, R.A. Olshen, and C.G. Stone. 1984. Classification and regression trees. The Wadsworth statistics/probability series. Chapman and Hall, New York, NY.
- Buehler, D.A., M.J. Welton, and T.A. Beachy. 2006. Predicting cerulean warbler habitat use in the Cumberland Mountains of Tennessee. *Journal of Wildlife Management* 70:1763-1769.
- Congalton, R.G. 1991. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of the Environment* 37:35-56.
- De'ath, G., K.A. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.
- Dettmers, R., and J. Bart. 1999. A GIS modeling method applied to predicting forest songbird habitat. *Ecological Applications* 9:152-163.
- Divine, D.D., D.W. Ebert, and C.L. Douglas. 2000. Examining desert bighorn habitat using 30-m and 100-m elevation data. *Wildlife Society Bulletin* 28:986-992.
- ESRI. 2006. ArcGIS Version 9.1. 380 New York St, Redlands, CA 92373-8100.

- Franklin, J. 1998. Predicting the distribution of shrub species in Southern California from climate and terrain-derived variables. *Journal of Vegetation Science* 9:733-748.
- Hahn, D.C., and R.J. O'Connor. 2002. Contrasting determinants of abundance in ancestral and colonized ranges of an invasive brood parasite. Pages 219-228 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, D.C., USA.
- Hamel, P., D.K. Dawson, and P.D. Keyser. 2004. How we can learn more about the cerulean warbler (*Dendroica cerulea*) *Auk* 121:7-14.
- Homer C, C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Land-cover Database for the United States. *Photogrammetric Engineering and Remote Sensing* 70:829-840.
- O'Brien, C.S., S.S. Rosenstock, J.J. Herver, J.L. Bright, and S.R. Boe. 2005. Landscape-level models of potential habitat for Sonoran pronghorn. *Wildlife Society Bulletin* 33:24-34.
- O'Conner, R.J., M.T. Jones, D. White, C. Hunsaker, T. Loveland, B. Jones, and E. Preston. 1996. Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States. *Biodiversity Letters* 3:97-110.
- Ontario Ministry of Natural Resources. 2008. Patch Analyst v4.0. <http://flash.lakeheadu.ca/~rrempel/patch/index.html>.
- Perkins, K.A. 2006. Cerulean Warbler selection of forest canopy gaps. MS Thesis, West Virginia University, Morgantown. <https://etd.wvu.edu/etd/documentdata.eTD?documentid=4596>
- R. 2008. The R Foundation for Statistical Computing. Version 2.6.2.
- Robbins, C.S., J.W. Fitzpatrick, and P.B. Hamel. 1992. A warbler in trouble: *Dendroica cerulea*. Pages 549-562 in J.M. Hagan III, and D.W. Johnston, editors. *Ecology and Conservation of Neotropical Migrant Landbirds*. Smithsonian Institution Press, Washington D.C.
- Rosenberg, K.V., and P.J. Blancher. 2005. Setting numerical population objectives for priority landbird species. Pages 57-67 C. J. Ralph and T. D. Rich, editors. *Bird Conservation Implementation and Integration in the Americas: Proceedings of the Third International Partners in Flight Conference*. U.S. Department of Agriculture Forest Service, General Technical Report PSW-GTR-191.
- Sauer, J.R., J.E. Fallon, and R. Johnson. 2003. Use of North American Breeding Bird Survey data to estimate population change for bird conservation regions. *The Journal of Wildlife Management* 67:372-389.
- Sauer, J. R., J. E. Hines, and J. Fallon. 2008. The North American Breeding Bird Survey,

- results and analysis 1966 - 2007. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, MD <http://www.mbr-pwrc.usgs.gov/bbs/bbs.html>.
- Scott, J.M., J.H. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson, editors. 2002. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, D.C., USA.
- Thogmartin, W.E., J.R. Sauer, and M.G. Knutson. 2004. A hierarchical spatial model of avian abundance with application to cerulean warblers. *Ecological Applications*. 14:1766-1779.
- Tobalske, C. 2002. Effects of spatial scale on the predictive ability of habitat models for the Green Woodpecker in Switzerland. Pages 197-204 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, D.C., USA.
- Vayssières, M.P., R.E. Plant, and B.H. Allen-Diaz. 2000. Classification trees: An alternative approach for predicting species distributions. *Journal of Vegetation Science*. 11:679-694.
- Weakland, C.A., and P.B. Wood. 2005. Cerulean Warbler (*Dendroica cerulea*) microhabitat and landscape-level habitat characteristics in southern West Virginia. *Auk* 122:497-508.
- Webster, C.R., and M.A. Jenkins. 2005. Course woody debris dynamics in the Southern Appalachians as affected by topographic position and anthropogenic disturbance history. *Forest Ecology and Management* 217:319-330.
- Wilds, S.P.V. 1996. Gradient analysis of the distribution of flowering dogwood (*Cornus florida*) and dogwood anthracnose (*Discula destructiva*) in western Great Smoky Mountains National Park. MS Thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC.
- Wood, P.B., S.B. Bosworth, and R. Dettmers. 2006a. Cerulean Warbler abundance relative to large-scale edge and habitat characteristics. *Condor*. 108:154-165.
- Wood, P.B., M.P. Strager, J.M. Strager. 2006b. Fine-scale forest bird habitat modeling for the mountaintop mining region within the Appalachian Bird Conservation Region. Final Report to the USGS, August 2006.
- Zabel, C.J., L.M. Roberts, B.S. Mulder, H.B. Stauffer, J.R. Dunk, K. Wolcott, D. Solis, M. Gertsch, B. Woodbridge, A. Wright, G. Goldsmith, and C. Keckler. 2002. A collaborative approach in adaptive management at a large-landscape scale. Pages 241-253 in J.M. Scott, P.J. Heglund, and M.L. Morrison, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Press, Washington, D.C., USA.



**Table 1.** Variables used to predict presence/absence and abundance of cerulean warblers in the Appalachian Bird Conservation Region (BCR28). Variables were only used at scales that would be biologically relevant to the focal species. For models created at each scale, B = both resolution groups and S = stop-level models only.

Variable Description	Scale <sup>a</sup>			Source
	100 m	300 m	1000 m	
Elevation	B	B	B	USGS 30 m National Elevation Dataset
Slope	S	S		USGS 30 m National Elevation Dataset
Aspect - Beers transformation	S	S		USGS 30 m National Elevation Dataset
Relative slope position	S	S		USGS 30 m National Elevation Dataset
Percent forest in buffer	B	B	B	2001 National Land Cover Dataset
Max size forest patch intersecting buffer	B	B	B	2001 National Land Cover Dataset
PFF score (forest connectivity)		B	B	2001 National Land Cover Dataset
Interior forest		B	B	2001 National Land Cover Dataset
Percent developed in buffer	B	B	B	2001 National Land Cover Dataset
Moisture index	S			USGS 30 m National Elevation Dataset
Distance to forested streams			B	NHDplus (National Hydrography Dataset)
Latitude	B (scaleless)			Breeding Bird Survey
Longitude	B (scaleless)			Breeding Bird Survey

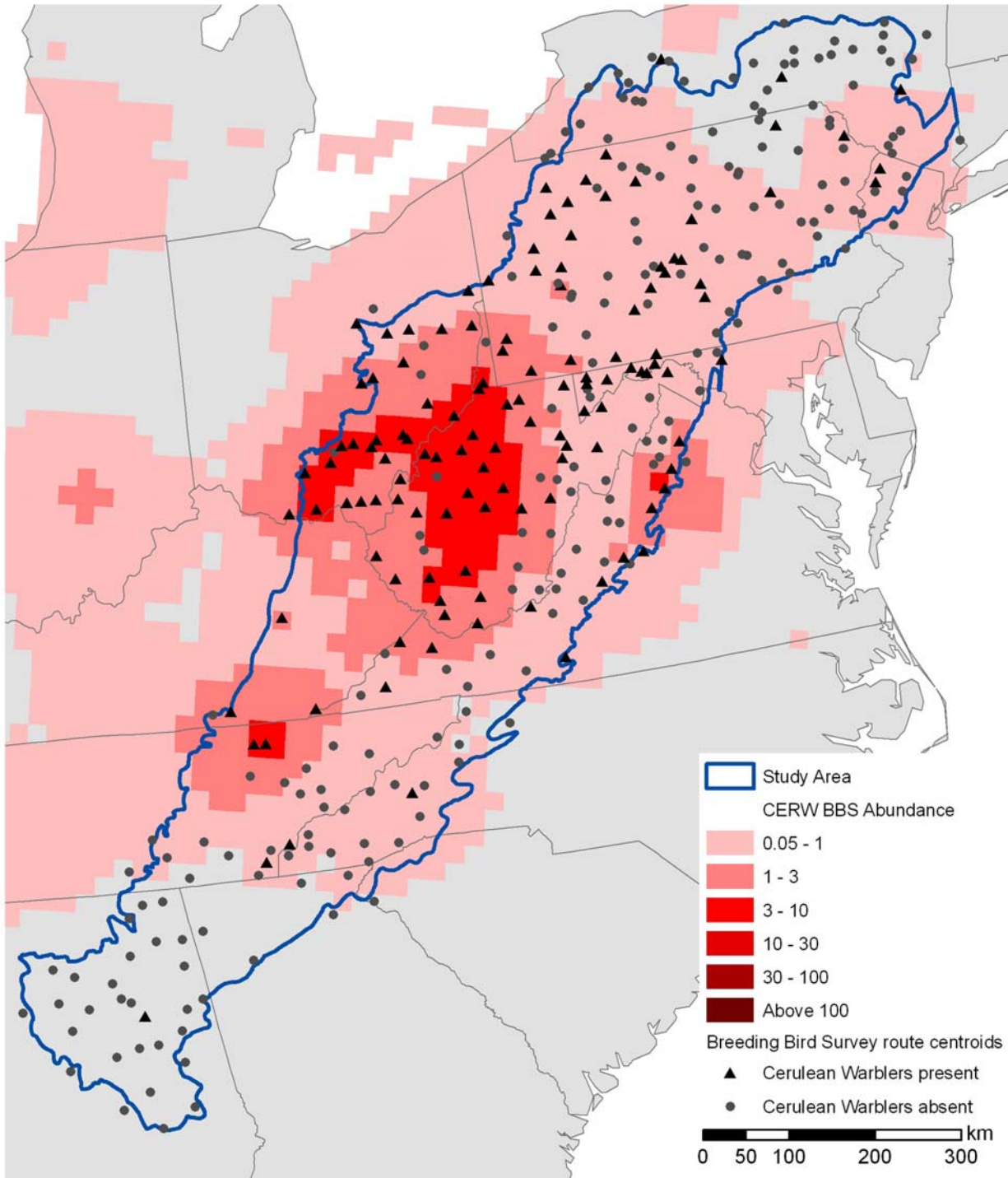
<sup>a</sup> Refers to the size of the buffer around each North American Breeding Bird Survey (NABBS) route.

**Table 2.** Confusion matrix from the route-level classification tree for cerulean warbler presence/absence. Randomly chosen NABBS routes (n=35) were withheld from model creation for evaluation.

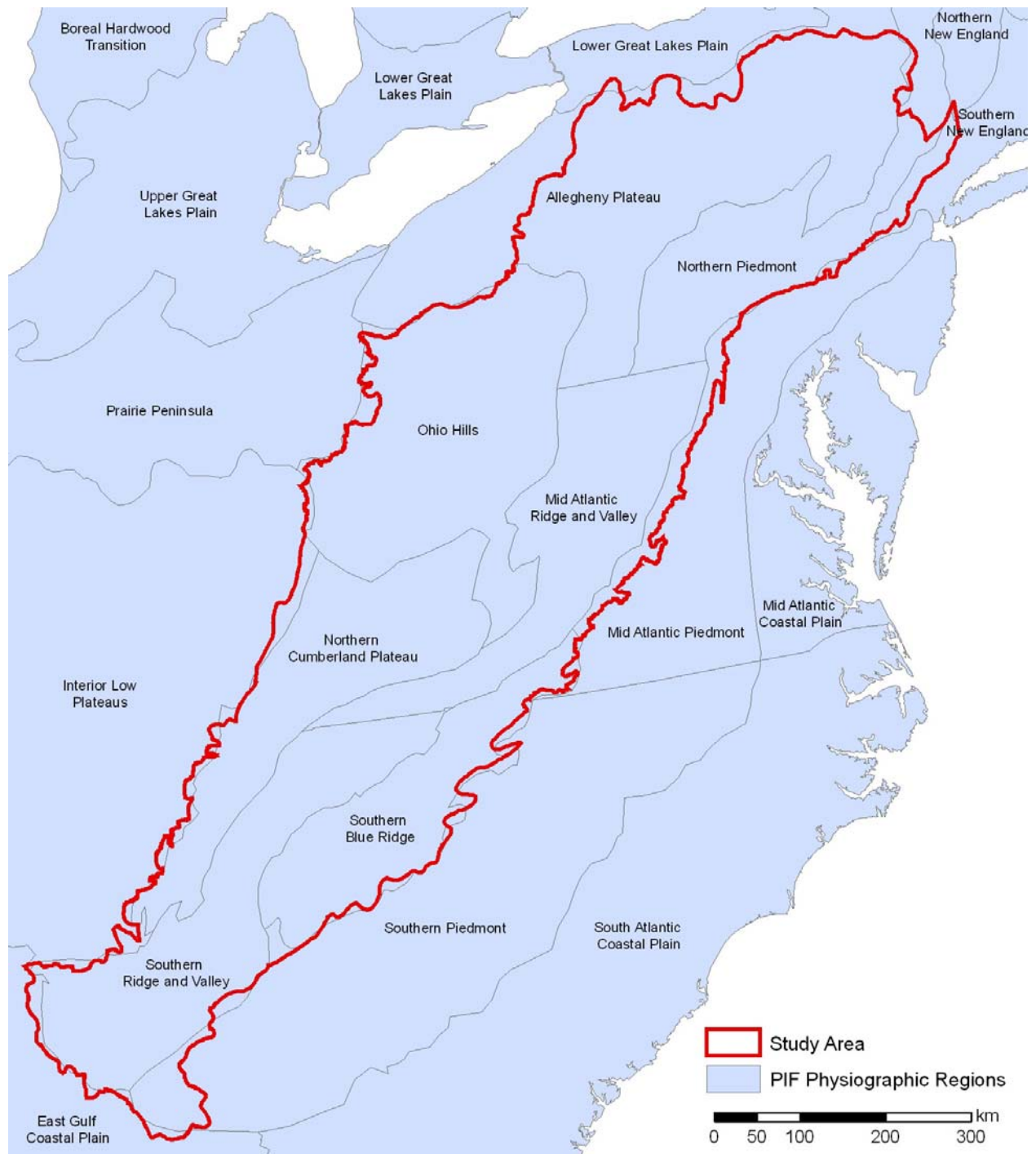
Actual Count Data			
Modeled Group	Absent	Present	Total
Absent	16	2	18
Present	6	11	17
Total	22	13	35
Accuracy	0.73	0.85	0.77
Kappa	0.54		

**Table 3.** Confusion matrix from the stop-level classification tree for cerulean warbler presence/absence. Randomly chosen NABBS stops (n=1607) were withheld from model creation for evaluation.

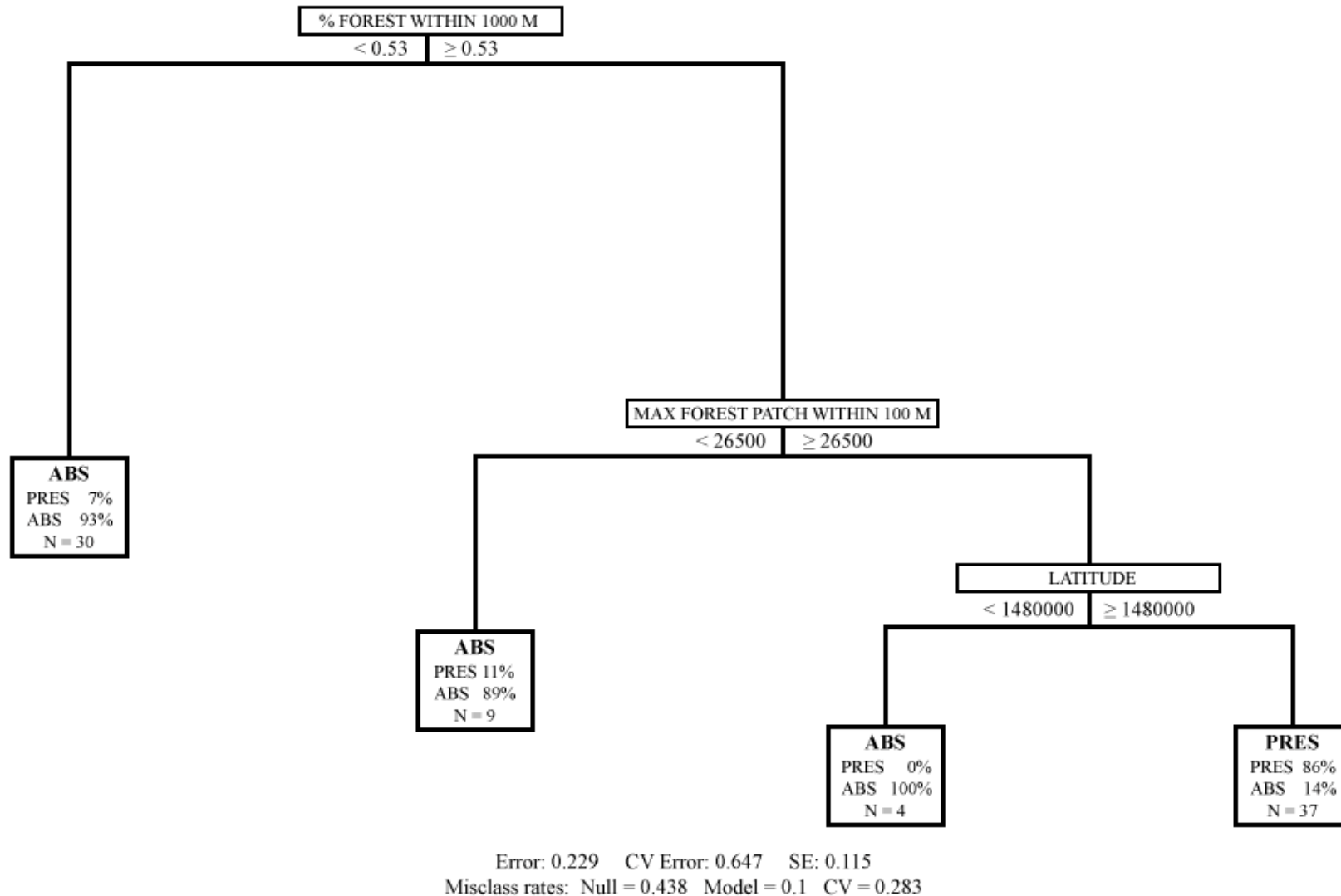
Actual Count Data			
Modeled Group	Absent	Present	Total
Absent	1509	75	1584
Present	10	13	23
Total	1519	88	1607
Accuracy	0.99	0.15	0.95
Kappa	0.22		



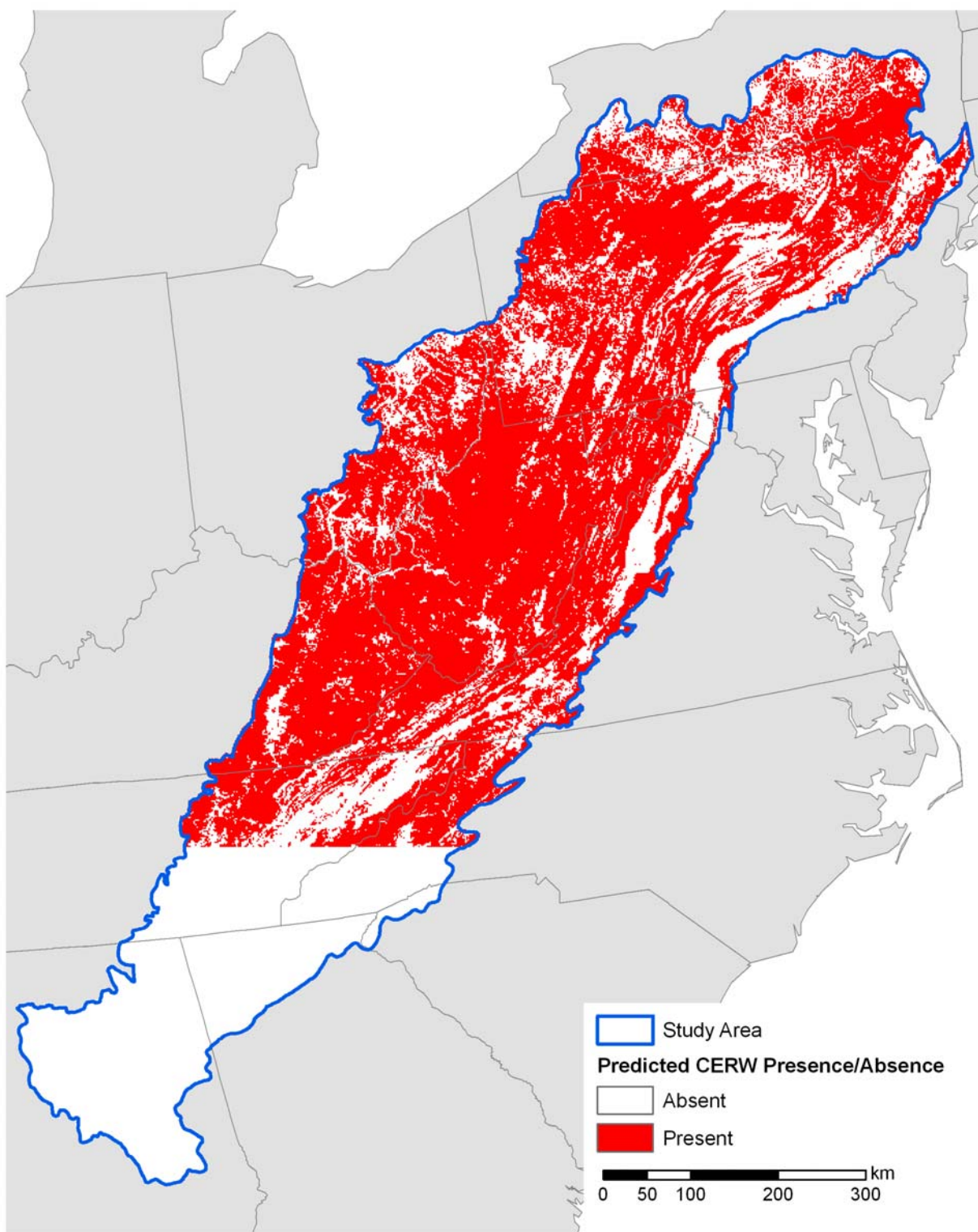
**Figure 1.** Breeding range of the cerulean warbler as derived from an interpolation of Breeding Bird Survey (NABBS) counts from 1961 – 2006 (Sauer et al. 2008). Black triangles represent centroids for NABBS routes where at least one Cerulean Warbler was detected between 1998 and 2004. Grey dots represent routes where no Cerulean Warblers were detected during the time period.



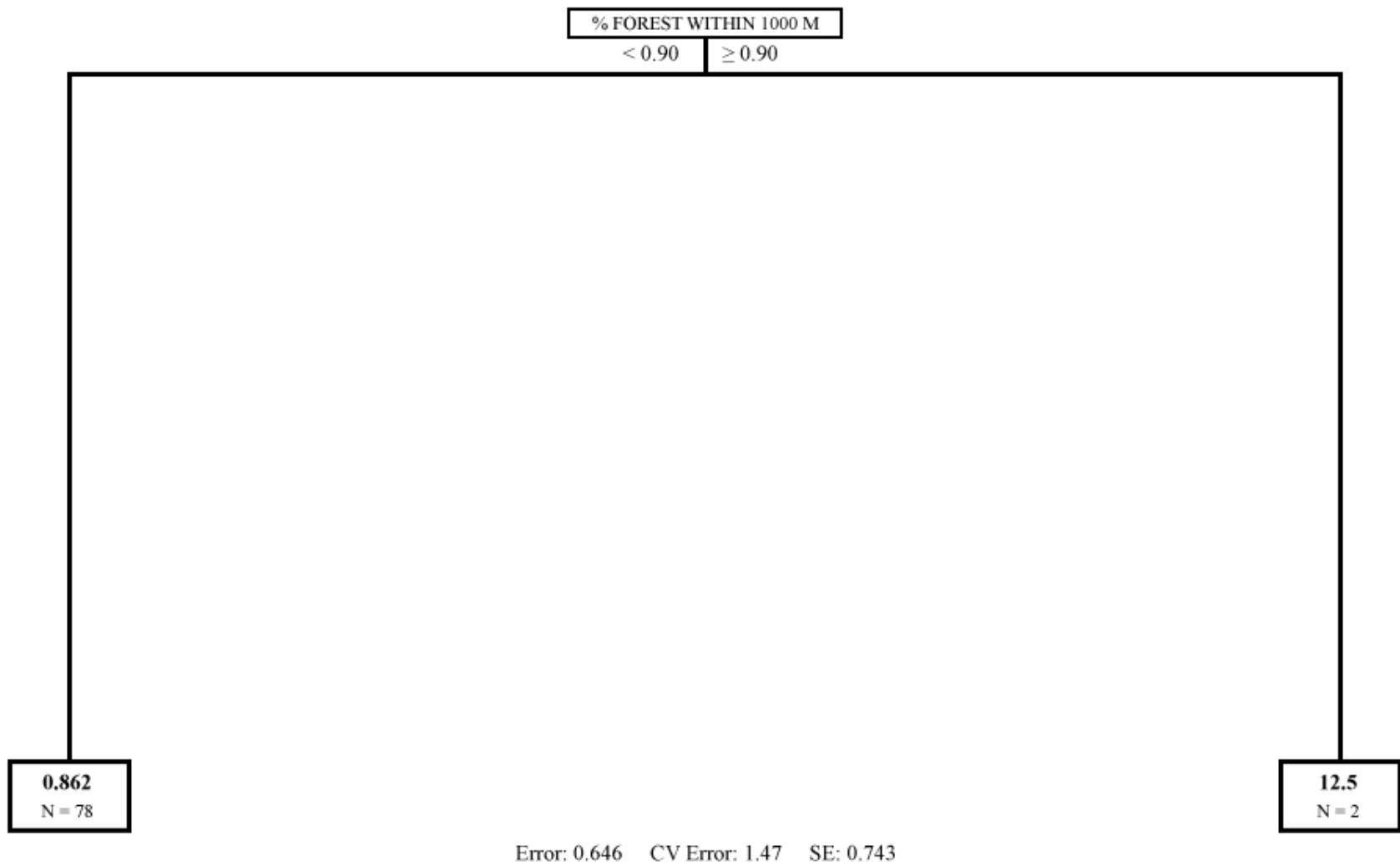
**Figure 2.** Partners in Flight physiographic regions and study area boundary.



**Figure 3.** Route-level classification tree model for cerulean warbler presence (PRES)/absence (ABS). Values below each node represent the number of routes assigned to each associated category.

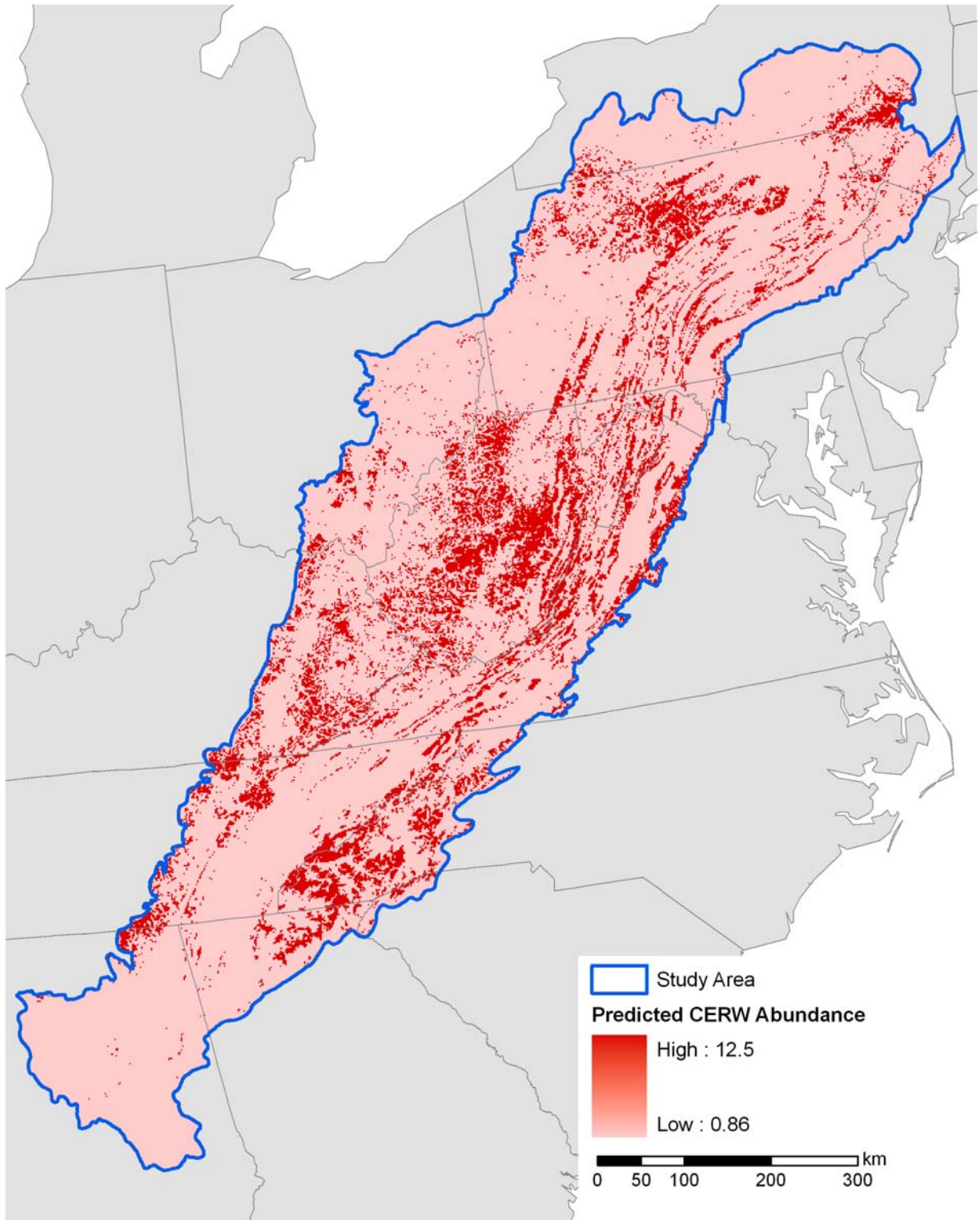


**Figure 4.** Predicted presence/absence of cerulean warblers within the Appalachian Bird Conservation region. Results are based on route-level classification tree analysis (n=115 NABBS routes).



**Figure 5.** Route-level regression tree model for cerulean warbler abundance. Each node ends with the estimated abundance of cerulean warblers and the number of routes classified by the node.

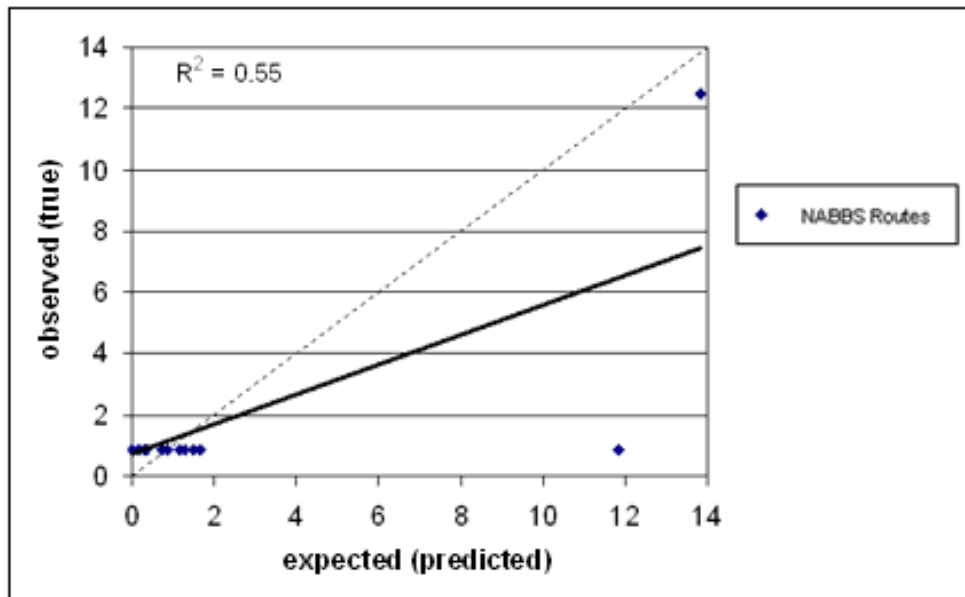




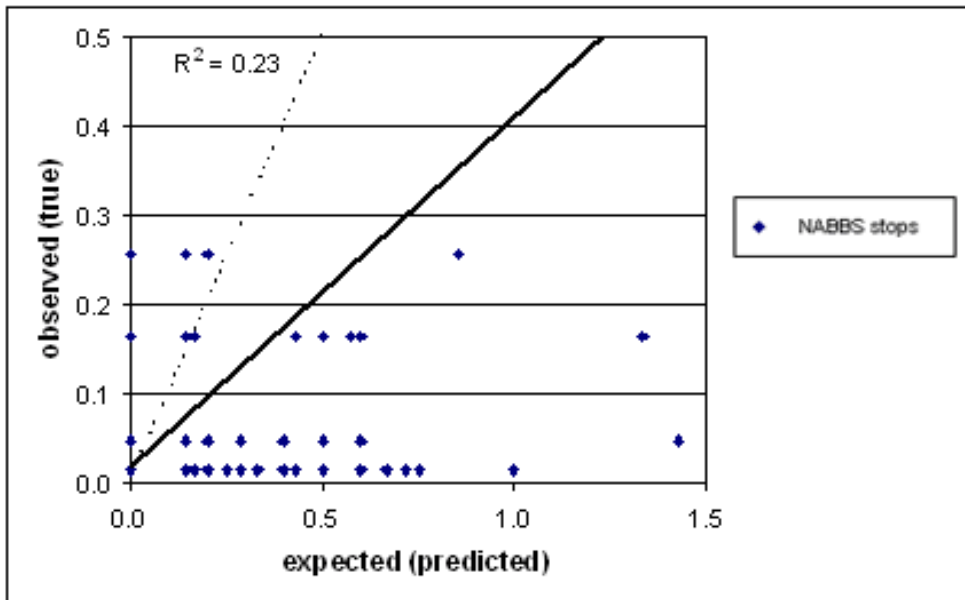
**Figure 6.** Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on route-level regression tree analysis. Abundance is defined by the sum of counts on a NABBS survey route (n=115 routes).



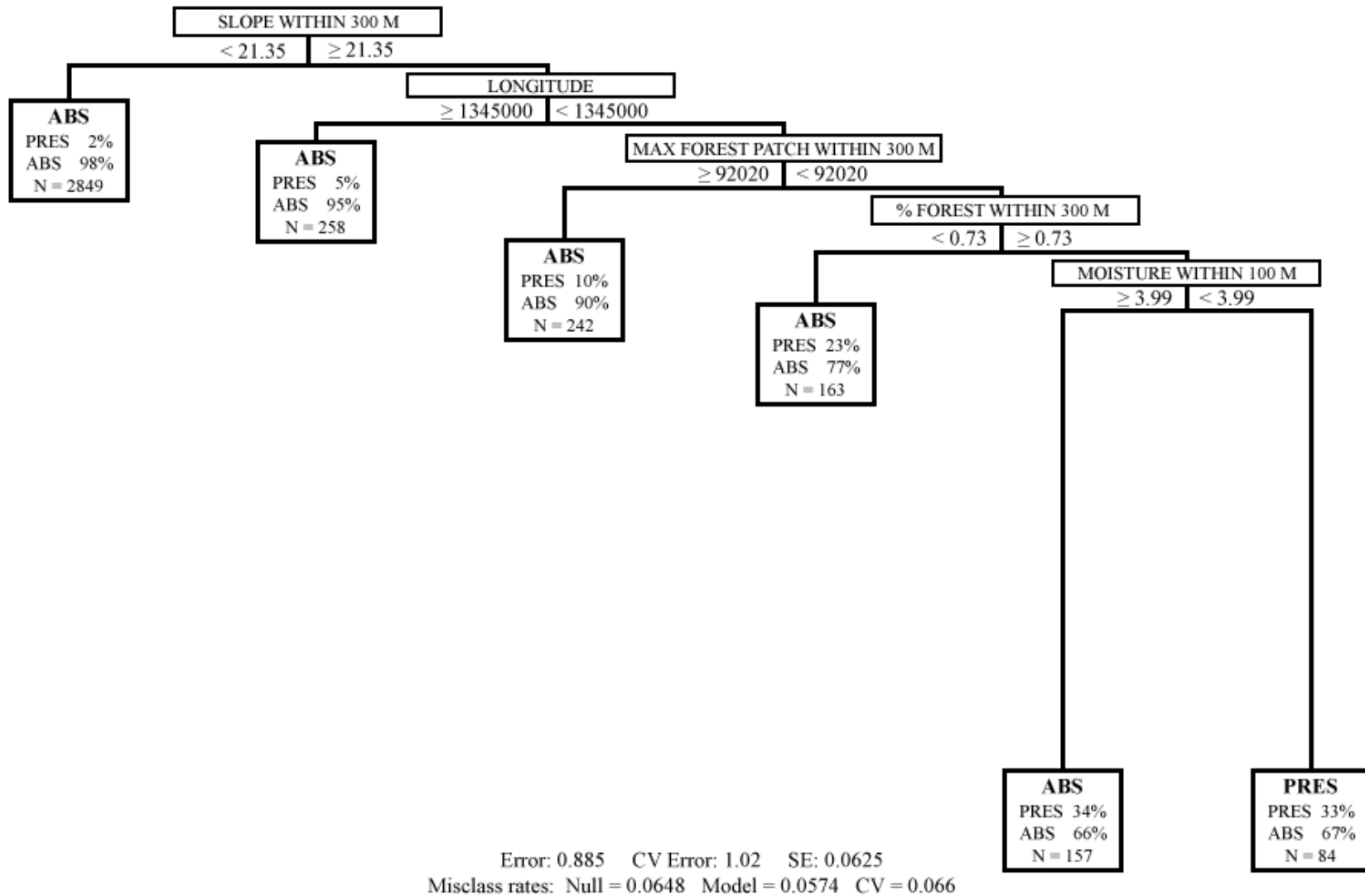
### a. Route-level regression tree model



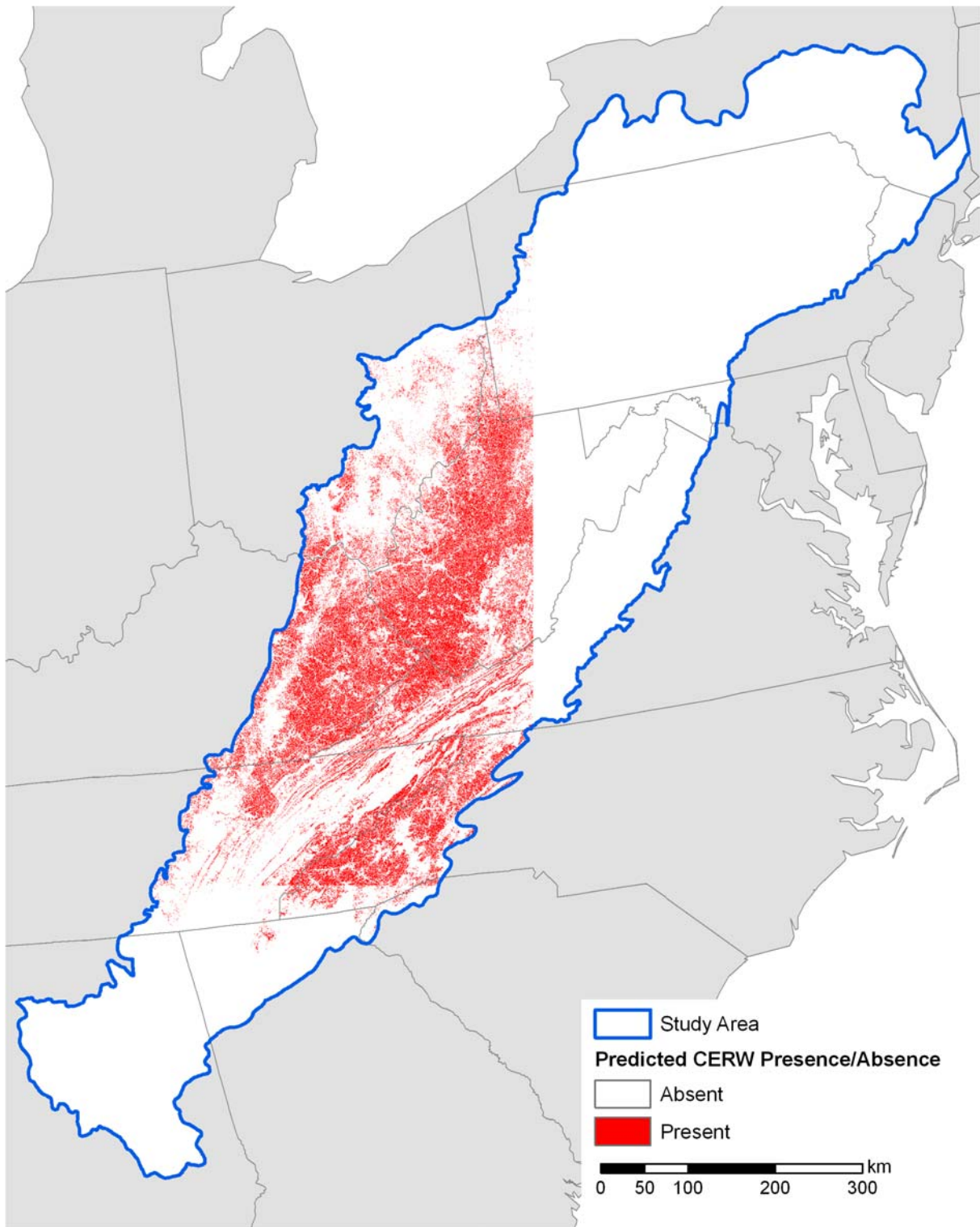
### b. Stop-level regression tree model



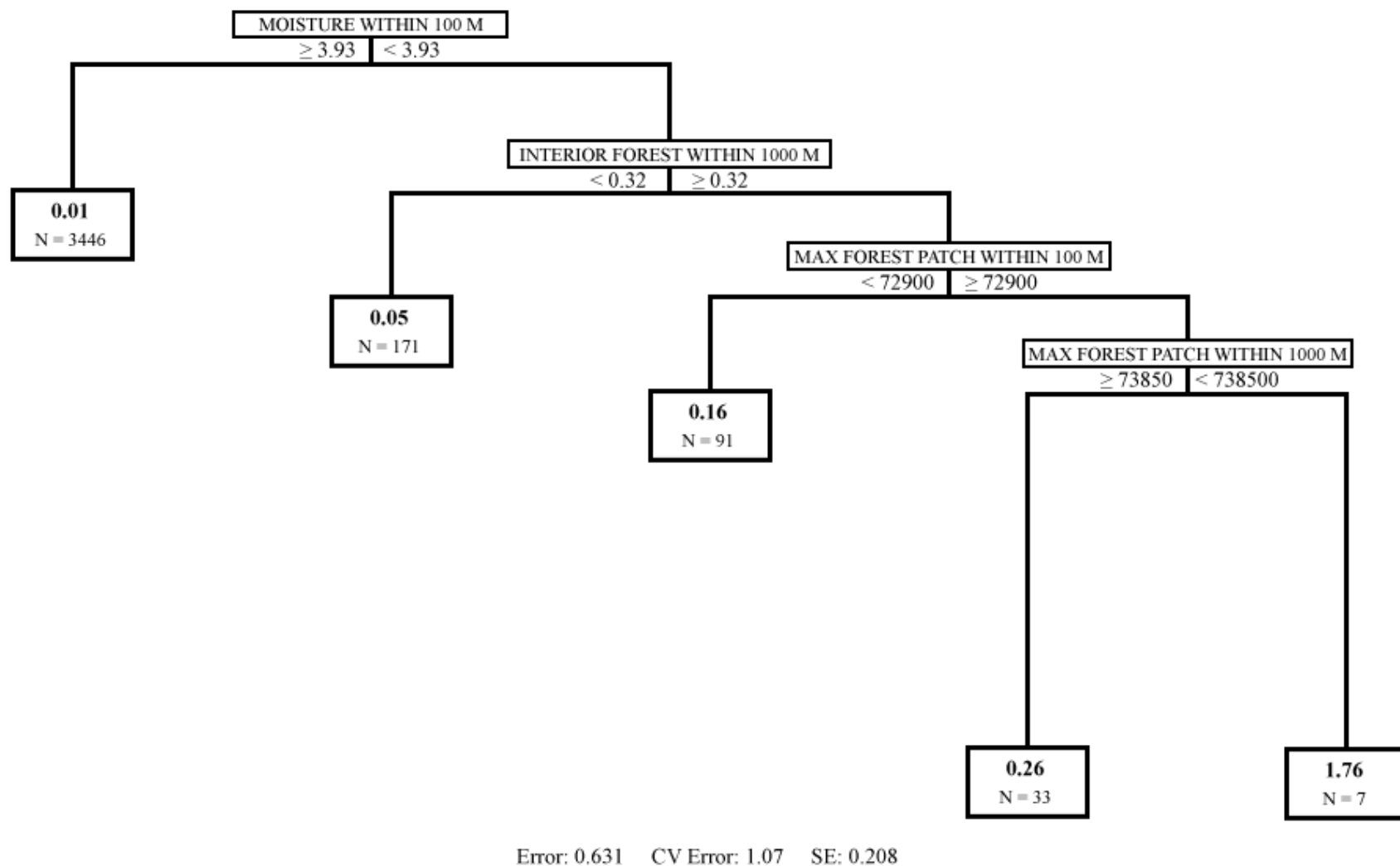
**Figure 7.** Observed mean North American Breeding Bird Survey (NABBS) counts (a:  $n = 35$  validation routes and b:  $n=1607$  validation stops) during 1998 – 2004 for cerulean warblers in the Appalachian Bird Conservation Region (BCR28) compared to expected (predicted) counts from the best regression tree. Observed counts were withheld from model construction. Dashed lines represent a one-to-one correspondence between observed and predicted counts. We fitted observed versus expected counts with simple linear regression (solid line). Each point may represent more than one route or stop.



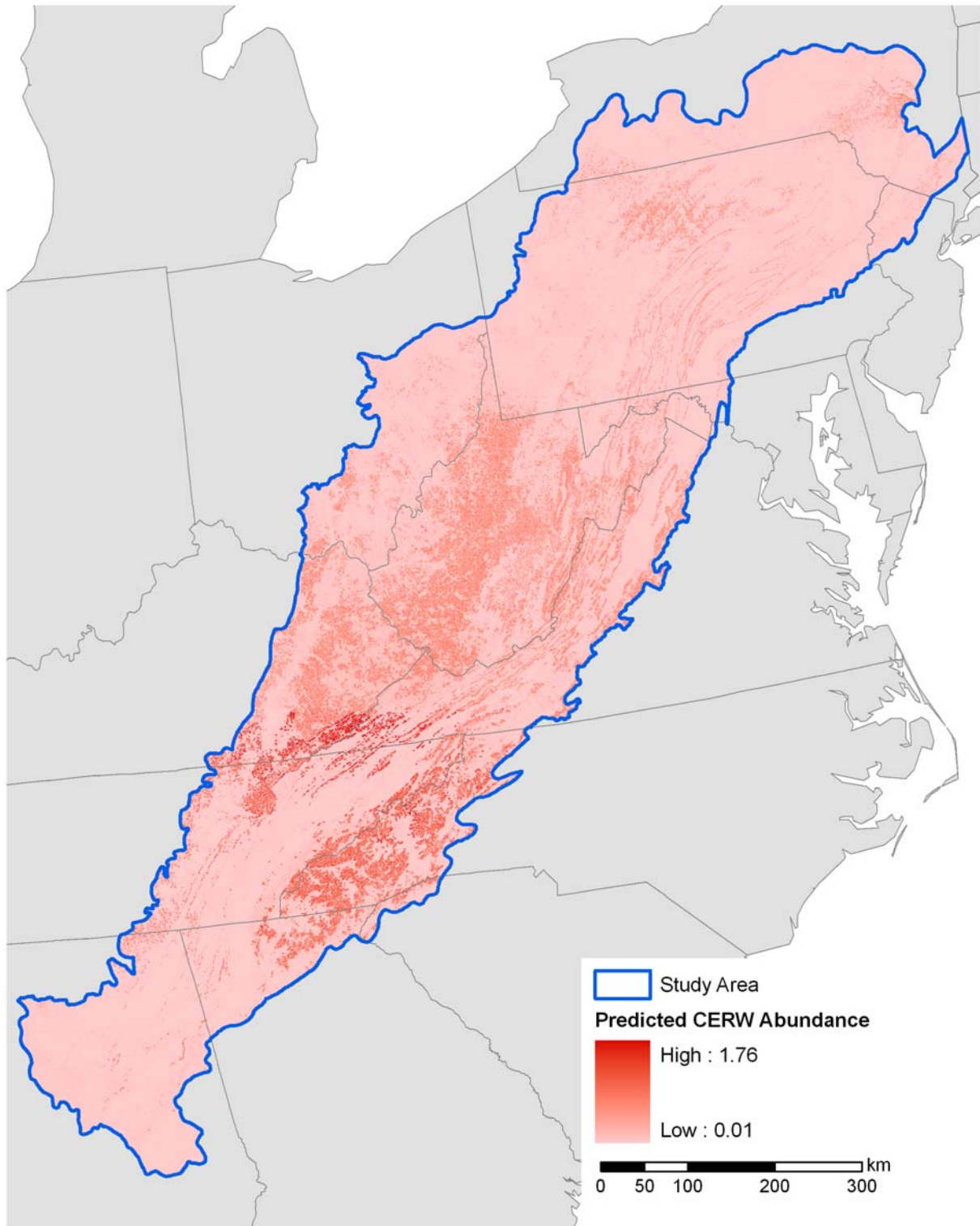
**Figure 8.** Stop-level classification tree model for cerulean warbler presence (PRES)/absence (ABS). Values below each node represent the number of routes assigned to each associated category.



**Figure 9.** Predicted presence/absence of cerulean warblers within the Appalachian Bird Conservation region. Results are based on stop-level classification tree analysis (n=5,355 NABBS stops).

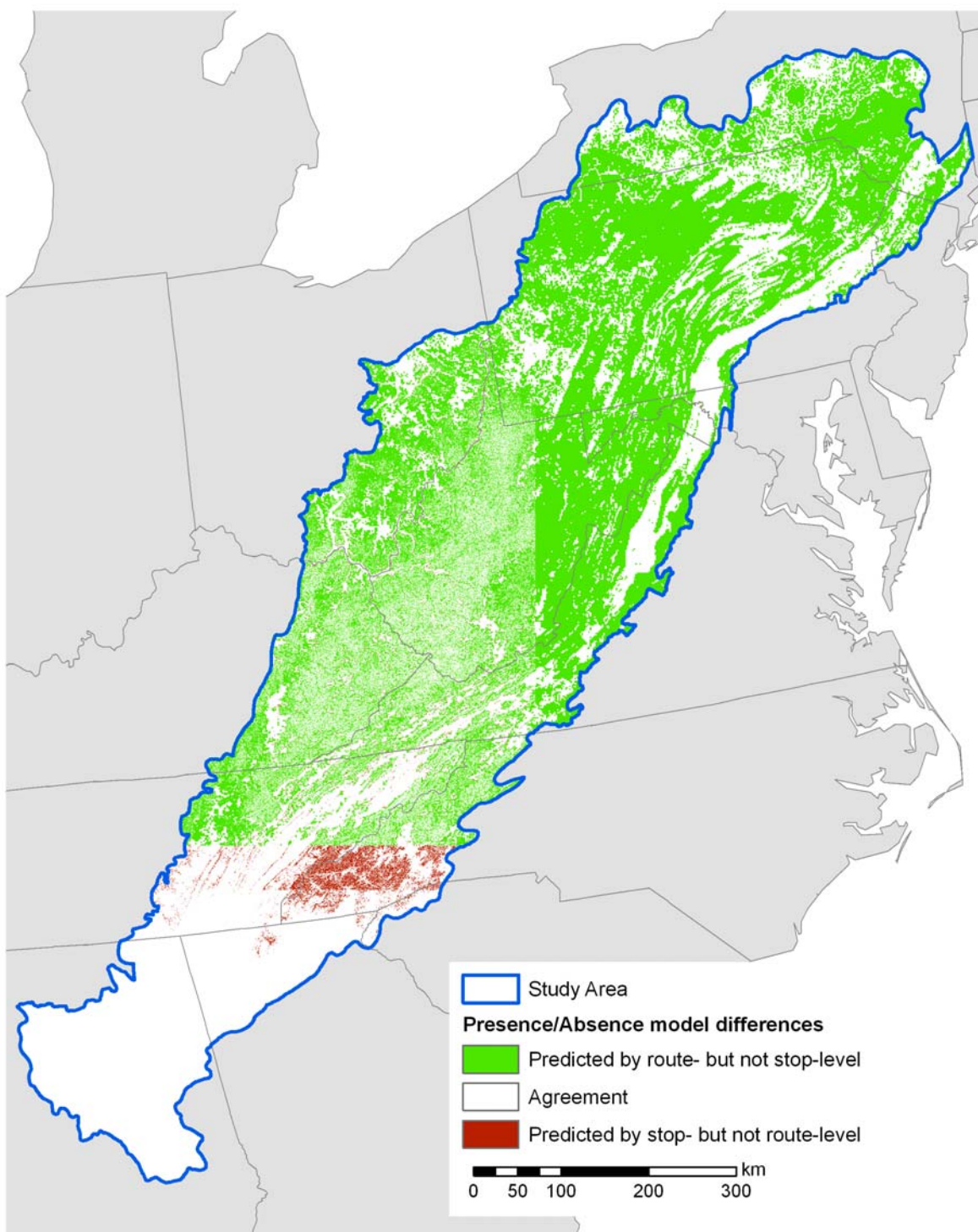


**Figure 10.** Stop-level regression tree model for cerulean warbler abundance. Each node ends with the estimated abundance of cerulean warblers and the number of routes classified by the node.

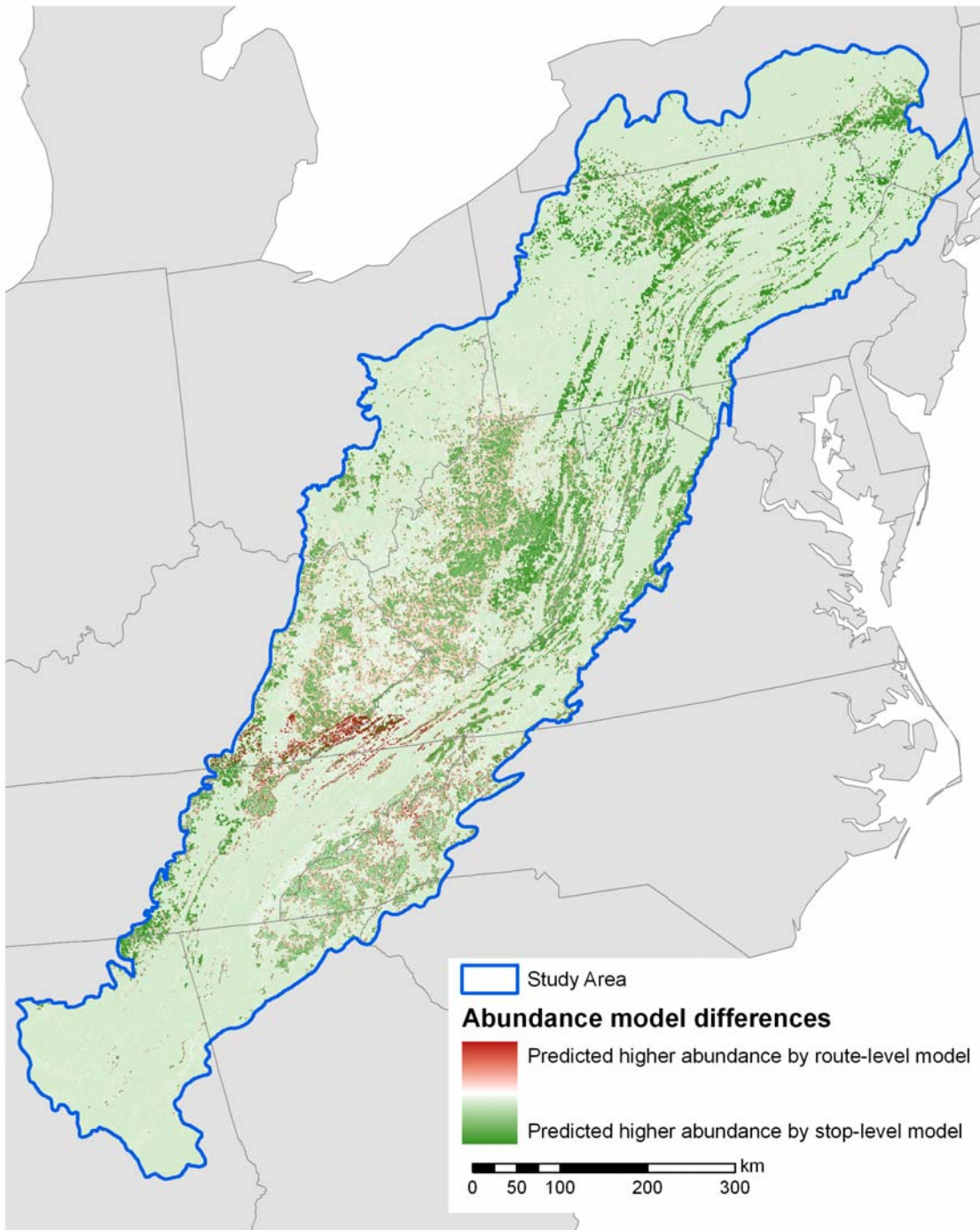


**Figure 11.** Predicted abundance of cerulean warblers within the Appalachian Bird Conservation region. Results are based on stop-level regression tree analysis (n=5,355 NABBS stops).



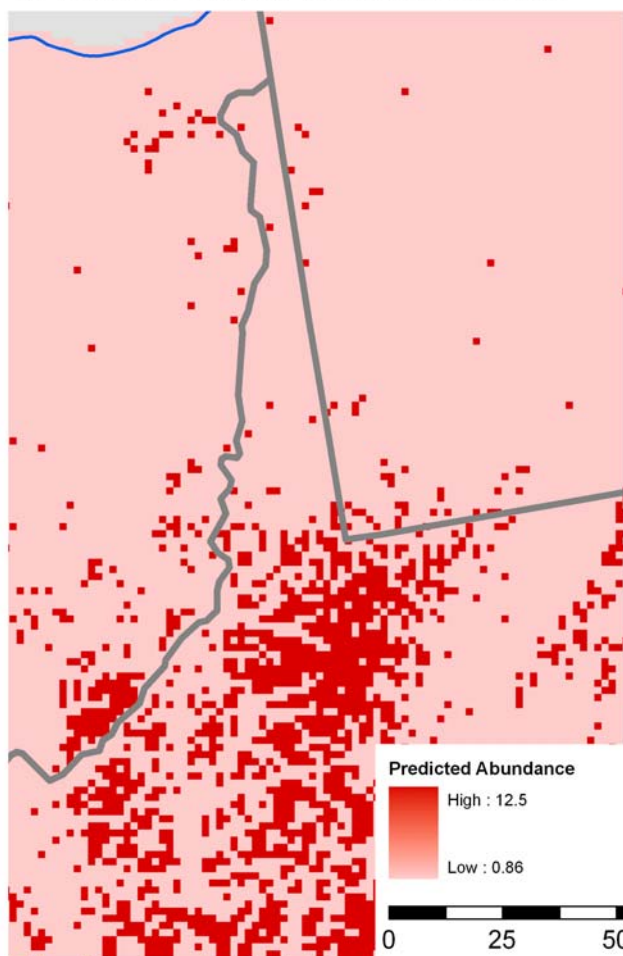


**Figure 12.** Variation between route- and stop-level maps for cerulean warbler presence/absence. Green cells indicate areas predicted to contain cerulean warblers by route- but not stop-level models. Red cells indicate areas predicted to contain cerulean warblers by stop- but not route-level models. Maps were created individually using classification trees and compared in ArcGIS 9.2.

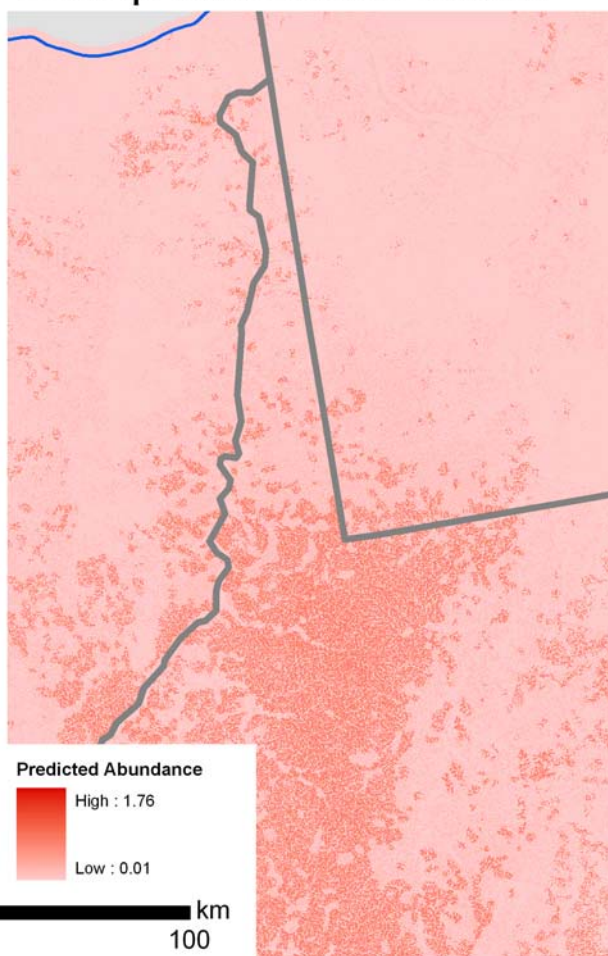


**Figure 13.** Variation between route- and stop-level maps for cerulean warbler relative abundance. Green cells indicate areas where cerulean warblers are predicted in higher abundances by route-level models. Red cells indicate areas where cerulean warblers are predicted in higher abundances by stop-level models. Maps were created individually using regression trees and compared by assessing the mathematical difference between raster cells in ArcGIS 9.2. Route-level maps were standardized by dividing abundances by 50 (number of NABBS stops on a route) for direct comparison to stop-level maps.

a. Route-level model results



b. Stop-level model results



**Figure 14.** Map resolution comparison between (a) route-level regression tree results and (b) stop-level regression tree results. Area represents the boundaries of Pennsylvania, West Virginia, and Ohio. Cell size for the route-level map was resampled to a cell size of 1566.5 m x 1566.5 m to represent the listening radius for cerulean warblers over 50 NABBS stops. Cell size for the stop-level map was 221.5 m x 221.5 m to represent the listening radius for cerulean warblers over a single NABBS stop.



## **Chapter 4**

### **Technological limitations of regional species modeling**

## INTRODUCTION

To date much ecological research has been limited in scope by available technology (Jager and Gross 2000). This is illustrated by many spatial analyses and ecological modeling projects. While the technology quickly advances in terms of data collection, there is still a lag for storage and processing technology. For example, it is currently possible to gather detailed information on canopy structure with an incredibly fine resolution using technology such as Light Detection and Ranging (LiDAR; Flood 2001, Queijja et al. 2005). However, not only can the costs of this type of product be extremely high, but the storage size required for such data is quite large. Additionally, the data are not easily integrated into spatial analyses and require a level of preprocessing before use. Aside from these issues of cost, storage space, and processing power, many of these products represent a snapshot in time and may not be applicable to the analysis of interest.

In this chapter I discuss some limitations of the analyses I conducted in chapters 2 and 3, highlight three major technological issues which complicate spatial analysis and ecological modeling (software limitations, hardware limitations, and mapping difficulties), and suggest research and technology solutions to address these needs and problems.

## SOFTWARE LIMITATIONS

While hardware is often a limiting factor for speed of analysis, limitations inherent in software may set a cap as to what can actually be analyzed, regardless of processing power. Many ecologists and land managers are using desktop personal computers (PCs) that are running a Microsoft® Windows 32-bit operating system. While operating systems such as Linux may use computer resources more efficiently, its use is relatively uncommon in ecology, and it may not be compatible with many software products commonly used by spatial ecologists (e.g., ArcGIS®, Erdas IMAGINE®). Here, I divide the modeling process into four stages: 1) survey data collection and organization, 2) derivation of landscape metrics, 3) data formatting for modeling, and 4) the modeling process (Appendices 1 and 2).

### *Survey data collection and organization*

Many modeling projects use datasets that have already been collected by other agencies or represent a long time series of data. Most of these datasets have traditionally been stored in

Microsoft® Excel spreadsheets. While the data can be stored in databases and executable files with a somewhat unlimited capacity, spreadsheet software such as Microsoft Excel has a limited capacity. Excel 2003 and previous versions could only display and process 65,535 rows and 256 columns (Microsoft 2008). The most recent version of the software features an increase in the limits: 16,000 columns and 1 million rows; but Excel has removed the ability to work with .DBF files, which are standard for many GIS applications. Traditionally these row and column limits may have been more than sufficient for most ecological datasets, but as landscape level models become more common, these limits will become problematic. Datasets covering a large geographic area (e.g., countries, physiographic regions) and time series data may easily exceed these limits. As an example, the Breeding Bird Survey (BBS) raw, unsummarized data are available as executable databases. Because of the scale of the BBS, the size of the databases exceeds the limits for use in Microsoft® Excel. While the data may be extracted in sections using an application such as SAS (SAS Institute 2003), the process can be cumbersome, tedious, and limit the analysis design. Row and column limitations are inherent throughout the analytical process. Additionally, SAS processes a limited number of columns; however, the number of rows in a dataset is not problematic.

#### *Derivation of landscape metrics*

As the survey data are being organized, various landscape coverages can be created based on the study area of interest. Typically this is done using geographic information systems (GIS) software such as ArcMap (ESRI 2008). Large study areas such as Bird Conservation Regions (BCRs, <http://www.abcbirds.org/abcprograms/domestic/landscape/BCR/>) often exceed the area (represented by number of pixels) that can be processed at one time. For instance, the Appalachian Mountains BCR is comprised of > 485,000,000 cells (30 m resolution). ArcMap has a limit of 256 unique values in a raster and a maximum of 65,536 attribute table entries for both raster and vector data coverages (ESRI 2008). Depending on the metric of interest and the size of the study area, the region may need to be subdivided into multiple sections to generate grids with a small enough range of values. If not done carefully, this creates artificial edge and introduces error into the dataset. For example, the calculation of most forest metrics (e.g., forest interior and PFF score) used in chapters 2 and 3 required the study region to be subdivided into 5 overlapping regions to generate grids and zonal statistics, which were later combined for analysis.

### *Data formatting for modeling*

Once the survey data have been organized and landscape metrics have been derived, the data are merged and formatted for modeling. Depending on the analysis type, this process may vary from simple to tedious and lengthy. Preparation for the decision tree analyses (De'ath and Fabricius 2000) in chapter 2 and 3 was simplistic; however, the formatting for the hierarchical spatial models of chapter 2 was complicated. WinBUGS (Spielgelhalter et al. 2003) requires the data to be formatted so that the values for random effects are numbered beginning with 1 and increase consecutively. Additionally, if a spatial term is added to the model, the data must be arranged to coincide with the numbering system assigned during the creation of the neighborhood structure and matrices. This third stage of the modeling process is generally the least problematic in terms of computer limitations. However, it should be noted that the same limitations inherent in the initial data collection and organization are present during this stage of the process.

### *The modeling process*

The final analysis stage of the process can be the most difficult in terms of computer processing. Because of the iterative nature of the Markov chain Monte Carlo procedure, it is often difficult to fit the hierarchical spatial models to large datasets where the number of random effects is high. At this stage of the modeling, I ran into the most trouble fitting the hierarchical spatial models. Route-level models were completed with all model terms and converged after a run of 100,000 iterations. However, the addition of a stop term and stop-level neighborhood structure introduced a level of complexity into the model that WinBUGS was not able to process. None of the stop-level models were able to finish successfully with the addition of the spatial conditional autoregression term (CAR). Even without the spatial CAR term, the models often terminated early in the iteration process due to 'trap' and 'black-box' errors (Figure 1), which are largely incomprehensible. The models were created with random effects for stop (or route for route-level models), observer, and year. Imputing random effects for 5,811 NABBS stops and 100,000 iterations caused the models to terminate early in the iteration process due to lack of memory. Models with a smaller geographic extent (Ohio Hills PIF Region 22; Figure 2) resulted in the same errors as the complete model for the entire Bird Conservation Region.

These models were run using WinBUGS v1.4 on a Windows-based desktop computer with 4GB of RAM and a quad-core processor. The computer performance was monitored during

the iteration process of WinBUGS to test for resource utilization. The computer's central processing unit (CPU) was not at maximum capacity during any of the model runs. This indicates that any memory issues resulting in premature model termination (trap and black-box errors) were a function of the software, and not hardware limitations. Even without the spatial CAR term, the stop-level models often terminated early. This memory issue was partially overcome by not sampling the stop term until later in the iteration process. Though the parameter estimates may not have been precise, the overall results did not seem affected (Wayne Thogmartin, pers. comm.). Ultimately, the stop-level hierarchical spatial models were replaced with a more simplistic decision tree analysis for chapter 3 because of issues with fitting the models and computing all the necessary results.

## **HARDWARE LIMITATIONS**

While software may be a major limiting factor for design and implementation of the models, hardware limitations can exacerbate the issue. Since most analyses are run on a desktop PC running a Microsoft® Windows XP or Vista 32-bit operating system, the computer is limited to a maximum of 4GB of memory (of which 3.5 is accessible). Even with the best current processor (currently, Intel® quad-core), models with an iterative process are often cumbersome and time-consuming. Older computers, and those with less memory and inferior processing units, may encounter errors more frequently during the iteration sequence.

Interactions between software and hardware must also be taken into consideration. While upgrading to 64-bit processing may be a potential solution, many software packages are not written to access 64-bit. Thus, the upgrade in hardware would not improve software performance. As 64-bit processing becomes more common, the ability of software to maximize this processing power should also become more common. Further, software written for multi-threaded computation may help to overcome memory issues. Multi-threaded computation has been implemented for research involving DNA sequencing (Guo et al. 2008), but has not received much attention for ecological analyses.

Depending on the system configuration, hard drive space may also be an issue for modeling projects with a large study area. Many of the metrics used in my analyses were in floating point raster form. Each raster was approximately 6 gigabytes (GB) in size, and required several derivatives representing each scale of analysis. With a large set of metrics, layers can

easily exceed several terabytes (TB) of data. Most personal computers are currently unable to store this much information. This problem can be overcome with the use of a server; however, this may not be available for all research projects.

## **MODELING AND MAPPING LIMITATIONS**

The science products from the research I conducted that are potentially most useful for conservation are maps of predicted occurrence and abundance. However, a map of a statistical model is itself a model, a flawed realization of the initial statistics. Because of the static nature of predictive maps, creating a visual representation of a statistical model is not easily accomplished. One of the benefits of the hierarchical spatial models is the ability to include nuisance effects (e.g., stop, time of day, season) into the models. Yet it is not intuitive how to incorporate these nuisance effects into a map of predicted presence/absence or relative abundance.

Another issue is multi-model inference. Theoretically, ecological processes are occurring at multiple spatial scales. Variables for use in spatial analyses are often measured at multiple spatial scales; however, some may not be used by a species at all scales, or we may imperfectly measure the response of species at the scales we examine. For the hierarchical spatial models I created, individual models were built using variables at a common spatial scale. The models were then combined using a model-averaging approach (Burnham and Anderson 2002). Other studies have used a similar model averaging approach for multi-scale modeling (e.g. Thogmartin et al. 2006, Fearer et al. 2007). But this is only one possible technique for a multi-scale representation. Other techniques, such as creation of a mixed-scale “best” model, have been used to create a model representing multiple spatial scales (Thogmartin et al. 2004).

Data-mining may be a valuable tool for exploring data sets and uncovering relationships across multiple scales. For example, decision trees have been used with multi-scale ecological datasets to identify potentially important environmental variables for species (Vayssières et al. 2000, Bourg et al. 2005, Wood et al. 2006). Though, results of decision tree models must be interpreted with the caveat that they are largely exploratory in nature and only return estimates of correlation.

Mapping temporally variable environmental covariates (e.g., May vegetation height) is also a difficult task. Spatial layers may be absent that represent data collected by field

technicians. Interpolation is often required to extrapolate these data to other unsampled regions, and incorporating temporally variable environmental data in static predictive maps is unclear.

## **TECHNOLOGY AND RESEARCH NEEDS**

As landscape-level analyses and decision support systems become more common, the need for more powerful personal computers will become crucial to the success of modeling efforts. With the exception of storage capacity, advancements in hardware are of less concern than software advancements. Personal computers with large amounts of storage space running software that can process large quantities of data will be required as landscape level analyses become more common. Specifically, spreadsheet and database software should be able to accommodate more than the current limits of row and column numbers; Microsoft's newest version of Excel can now handle 16,000 columns and 1 million rows, for instance. Similarly, GIS applications such as ArcMap must be able to process large rasters with a wide range of unique cell values.

Making software more user-friendly will also aid in creating a greater number of models and more useful predictive models. WinBUGS, currently a low-generation version (1.4), is not user-friendly and is unable to take full advantage of the computer's processing resources. Related packages such as classic BUGS, and an open-source version of the Windows software, OpenBUGS, may overcome some of the memory issues associated with WinBUGS. However, these applications are generally customized and require a level of programming uncommon to many ecologists. Integration with other software packages, such as Excel and ArcMap, may extend the functionality of WinBUGS.

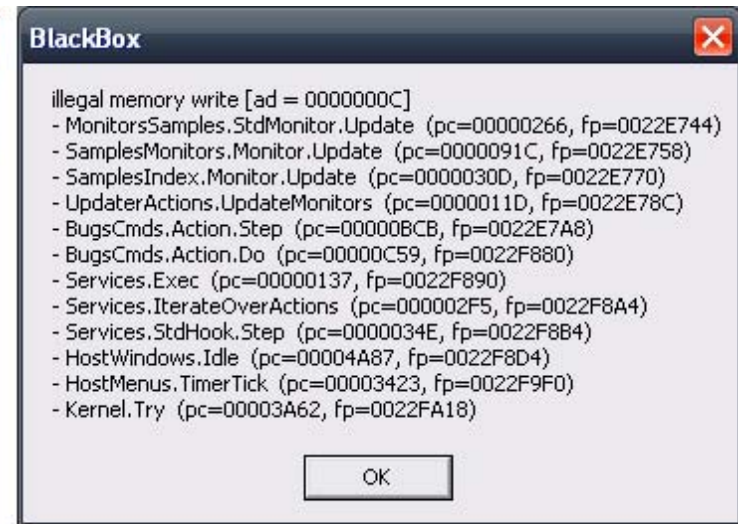
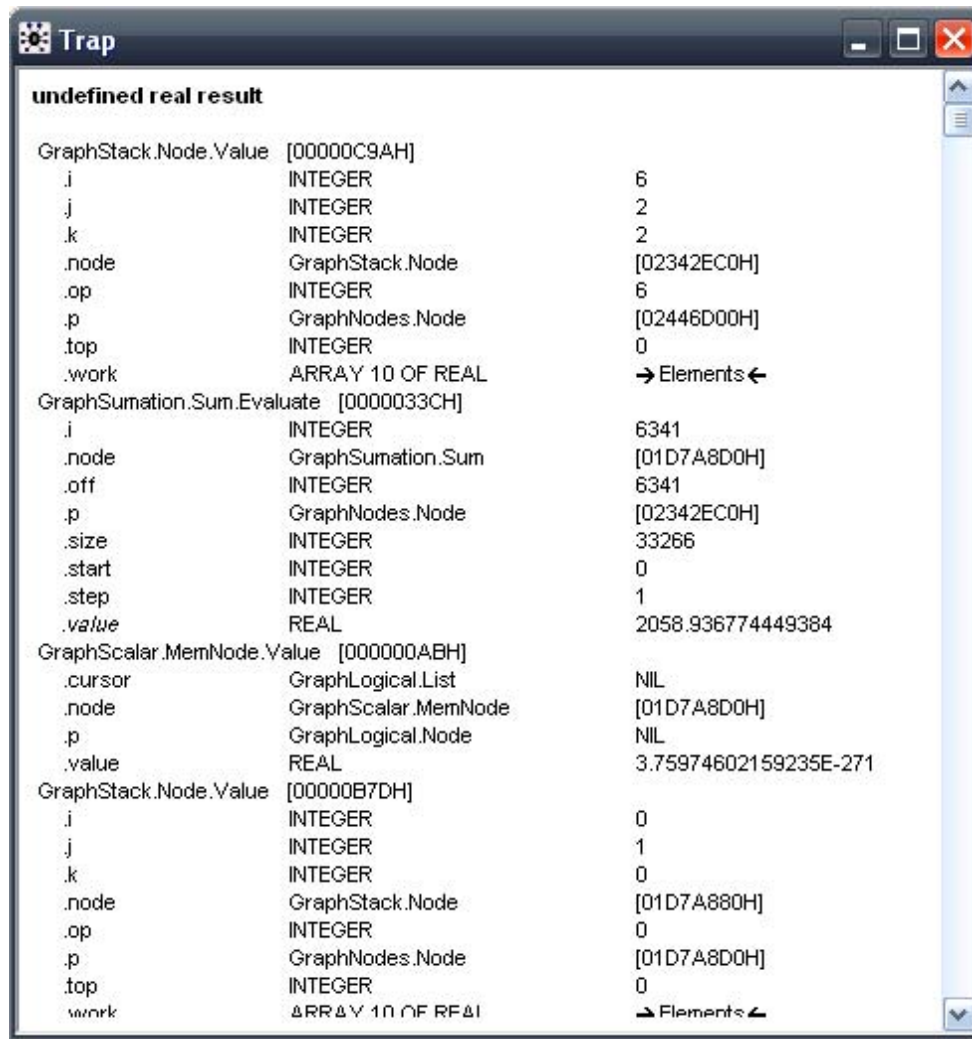
Overcoming technological resource limitations is only part of advancing ecological modeling. There are still many theoretical issues that need attention. Scale will always be at the forefront of these issues; we may never know at exactly which scales ecological phenomena occur or at how many scales. Additionally, it is not intuitive how to mix these scales into a single model. Future research needs to address these issues and find scales that may be the most relevant for a particular species or guild.

## LITERATURE CITED

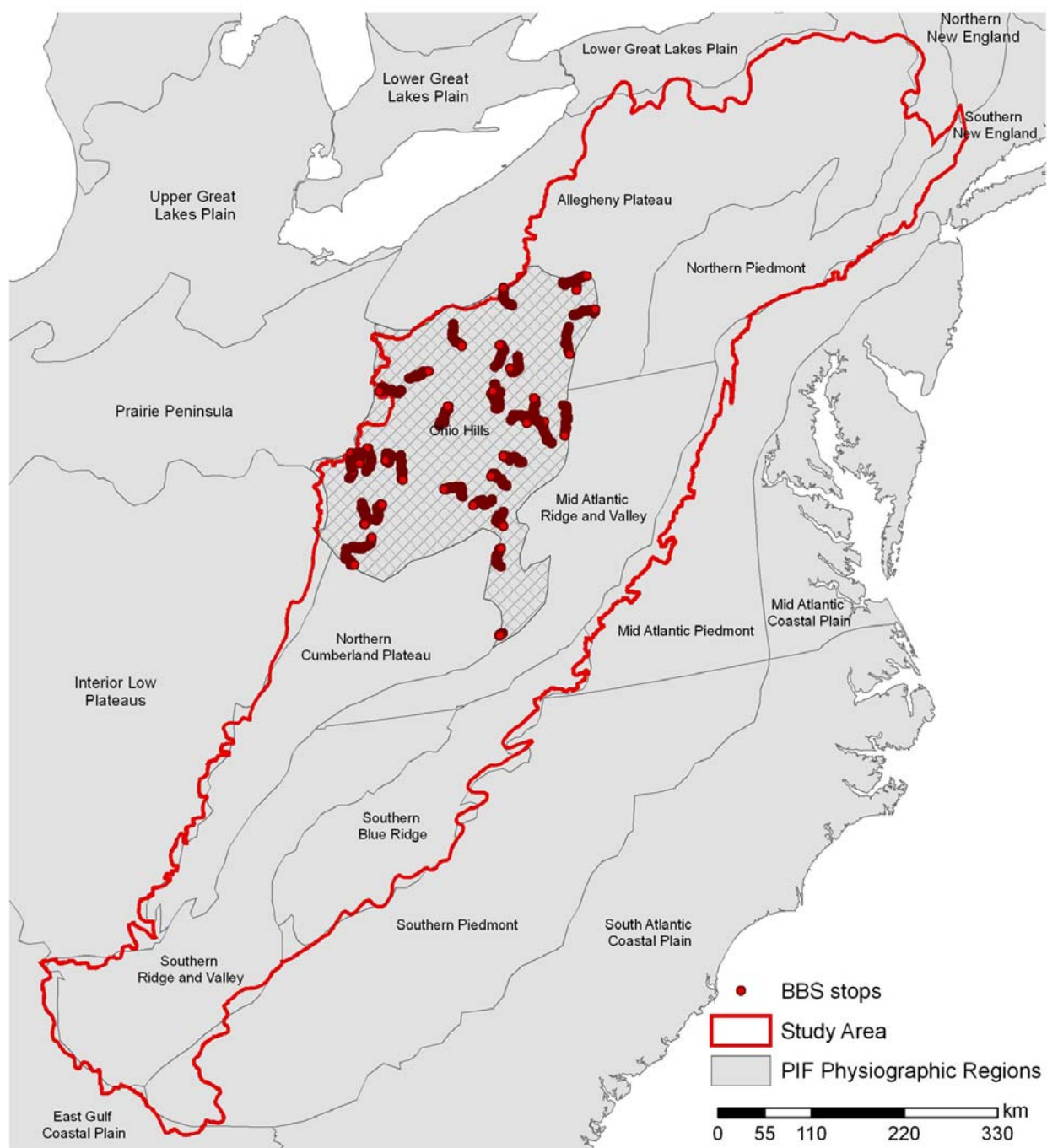
- Bourg, N.A., W.J. McShea, and D.E. Gill. 2005. Putting a CART before the search: successful habitat prediction for a rare forest herb. *Ecology* 86:2793-2804.
- Burnham, K.P., and D.R. Anderson. 1998. Model selection and inference: a practical information theoretic approach. Springer-Verlag, New York, NY.
- De'ath, G., and K.E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.
- ESRI. 2008. ArcGIS Version 9.2. 380 New York St, Redlands, CA 92373-8100.
- Fearer, T.M., S.P. Prisley, D.F. Stauffer, and P.D. Keyser. 2007. A method for integrating Breeding Bird Survey and Forest Inventory and Analysis databases to evaluate forest bird-habitat relationships at multiple spatial scales. *Forest Ecology and Management* 243:128-143.
- Flood, M. 2001. Laser altimetry: From science to commercial LiDAR mapping. *Photogrammetric Engineering and Remote Sensing* 67:1209-1217.
- Guo, T., L. Guiyang, and R. Deaton. 2008. Accelerating Computation of DNA Sequence Alignment in Distributed Environment. Volume 15 in *Advanced Intelligent Computing Theories and Applications With Aspects of Contemporary Intelligent Computing Techniques*. Springer Berlin Heidelberg. pp. 222-228.
- Jager, H.I., and L.J. Gross. 2000. Spatial Control: The Final Frontier in Applied Ecology. *Ecology* 81:1473-1474.
- Microsoft. 2008. Row height and column width limits in Excel. <http://office.microsoft.com/en-us/excel/HA101375451033.aspx>.
- Quieja, V.R., J.M. Stoker, and J.J. Kosovich. 2005. Recent U.S. Geological Survey applications of LiDAR 71:5-9.
- SAS Institute. 2003. Version 9.1. SAS Institute, Cary, North Carolina, USA.
- Spielgelhalter, D.J., A. Thomas, and N.G. Best. 2003. WinBUGS version 1.4 user manual. BRC Biostatistics Unit, Cambridge, UK. Available from <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.
- Thogmartin, W.E., J.R. Sauer, and M.G. Knutson. 2004. A hierarchical spatial model of avian abundance with application to cerulean warblers. *Ecological Applications* 14:1766-1779.
- Thogmartin, W.E., M.G. Knutson, and J.R. Sauer. 2006. Predicting regional abundance of rare grassland birds with a hierarchical spatial count model. *Condor* 108:25-46.



- Vayssières, M.P., R.E. Plant, and B.H. Allen-Diaz. 2000. Classification trees: An alternative approach for predicting species distributions. *Journal of Vegetation Science*. 11:679-694.
- Wood, P.B., M.P. Strager, J.M. Strager. 2006. Fine-scale forest bird habitat modeling for the mountaintop mining region within the Appalachian Bird Conservation Region. Final Report to the USGS, August 2006.

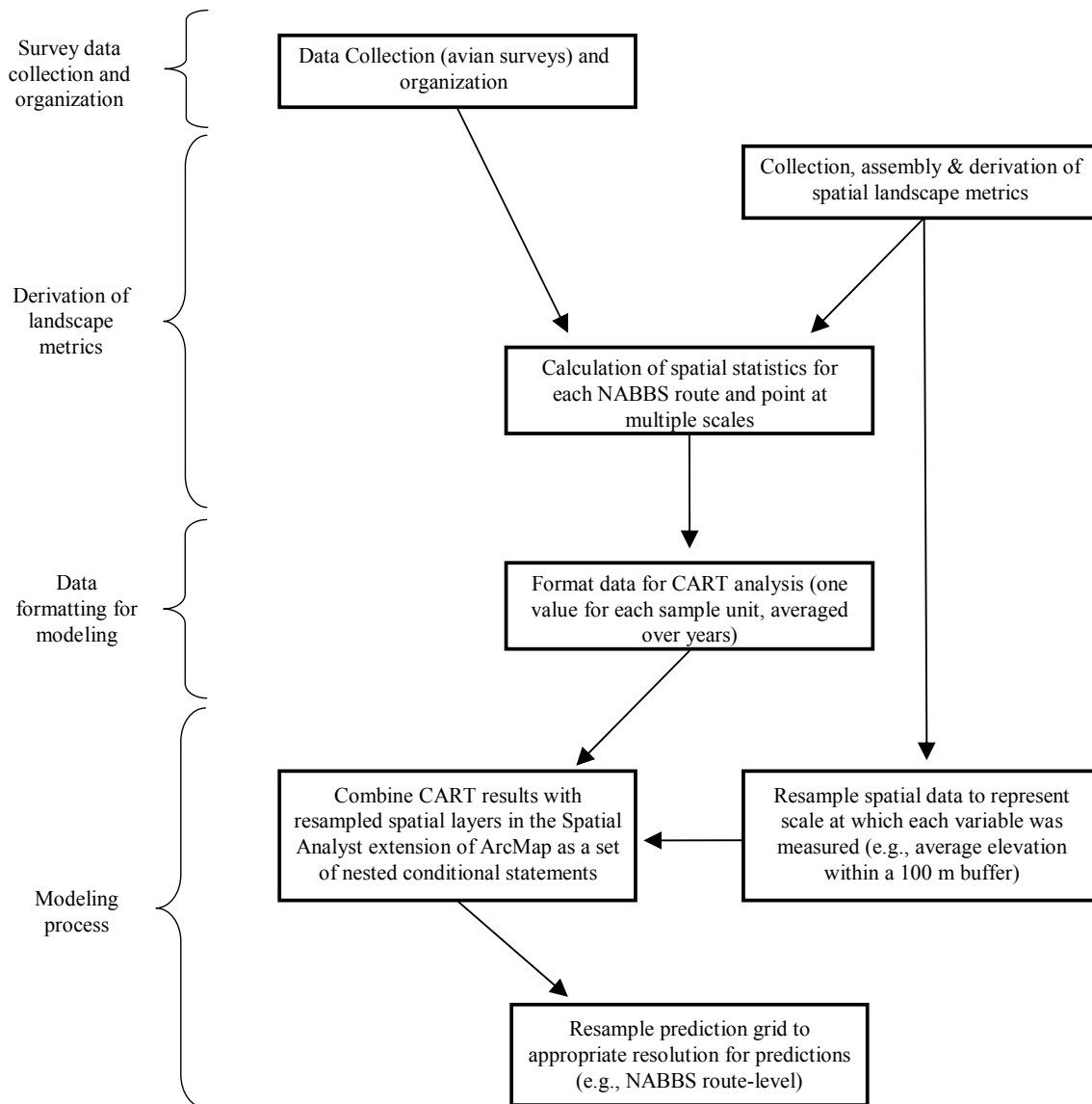


**Figure 1.** Examples of 'trap' and 'black-box' errors produced by WinBUGS during the model iteration.



**Figure 2.** Breeding Bird Survey stop locations within the Ohio Hills Partners in Flight physiographic region (PIF22).

**Appendix 1.** Complete process from data collection to model results for a classification and regression tree (CART) analysis.



**Appendix 2.** Complete process from data collection to model results for a hierarchical spatial analysis using WinBUGS.

